Characterisation of the CRISPR/Cas system of the hyperthermophilic Archaeum *Thermoproteus tenax*

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

der Fakultät für

Biologie und Geografie

an der

Universität Duisburg-Essen

vorgelegt von

ANDRÉ PLAGENS

aus Bottrop

Februar 2010

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Institut für Biologie in der Abteilung für Mikrobiologie I der Universität Duisburg-Essen, Campus Essen durchgeführt.

1. Gutachter: Prof. Dr. Reinhard Hensel (Essen)
2. Gutachter: Prof. Dr. Hemmo Meyer (Essen)
3. Gutachter: Prof. Dr. Reinhard Sterner (Regensburg)

Vorsitzende des Prüfungsausschusses: Prof. Dr. Shirley Knauer (Essen)

Tag der mündlichen Prüfung: 23.06.2010

Man will nicht nur glücklich sein,
sondern glücklicher als die anderen.
Und das ist deshalb so schwer,
weil wir die anderen für glücklicher halten,
als sie sind.

Charles-Louis de Montesquieu

TABLE OF CONTENTS

LIST OF FIGURES ..... **V**

LIST OF TABLES ..... **VI**

1 INTRODUCTION **1**

2 MATERIAL AND METHODS **9**

2.1 Chemicals, enzymes, kits and consumables **9**

2.2 Instruments **10**

2.3 Strains and growth conditions **12**

2.3.1 Culture conditions for growing *T. tenax* 12

2.3.1.1 Culture conditions at increased ionic strength 12

2.3.1.2 Culture conditions under UV-radiation 13

2.3.1.3 Culture conditions at sub- and supraoptimal temperatures 13

2.3.2 Culture conditions for *E. coli* 13

2.4 Plasmids and constructed recombinant vectors **14**

2.5 Working with RNA **15**

2.5.1 Treatment of solutions, glassware and equipment 15

2.5.2 Isolation of small and total RNA from *T. tenax* 15

2.5.2.1 Isopropanol precipitation 15

2.5.2.2 RNeasy Mini Kit (QIAGEN) 16

2.5.2.3 *mir*Vana™ miRNA Isolation Kit (Ambion) 16

2.5.3 Quantitative and qualitative analysis of RNA 17

2.5.4 Gel electrophoresis of RNA 17

2.5.4.1 Denaturing agarose gel electrophoresis 17

2.5.4.2 Denaturing polyacrylamide gel electrophoresis 18

2.5.5 Capillary transfer of RNA to a nylon membrane (Northern blot) 18

2.5.6 Generation of specific, DIG-labelled RNA probes by *in vitro* transcription 19

2.5.7 Hybridisation of immobilised total RNA with DIG-labelled probes 21

2.5.7.1 Hybridisation of total RNA with antisense RNA probes 21

2.5.7.2 Hybridisation of small RNA with antisense oligonucleotide probes 21

2.5.8 Immunological detection of RNA hybrids 22

2.5.9 Nuclease assay 22

2.5.10 Electrophoretic Mobility Shift Assay (EMSA) 23

2.5.11 cDNA synthesis of total RNA from *T. tenax* 23
2.6 Working with DNA 24
2.6.1 Preparation of genomic DNA from *T. tenax* 24
2.6.2 Preparation of plasmid DNA from *E. coli* 25
2.6.2.1 Preparation of plasmid DNA by alkaline lysis 25
2.6.2.2 Plasmid preparations with commercially available kits 25
2.6.2.3 Preparation of plasmid-DNA for colony PCR 26
2.6.3 DNA precipitation 26
2.6.4 Quantitative and qualitative analysis of DNA 26
2.6.5 Electrophoresis of DNA 26
2.6.5.1 Agarose gel electrophoresis of DNA 26
2.6.5.2 Non-denaturing polyacrylamide gel electrophoresis of DNA 27
2.6.6 Purification of DNA fragments 27
2.6.6.1 Gel extraction of agarose gels 27
2.6.6.2 Purification of PCR fragments 28
2.6.6.3 Electroelution from polyacrylamide gels 28
2.6.7 Polymerase chain reactions (PCR) 28
2.6.7.1 Amplification of genomic DNA and plasmid DNA 29
2.6.7.2 Amplification of primary cDNA 29
2.6.7.3 PCR mutagenesis 30
2.6.7.4 Overlap extension PCR 30
2.6.8 Enzymatic modification of DNA 31
2.6.8.1 Restriction of DNA 31
2.6.8.2 5'-dephosphorylation of linearised vector-DNA 31
2.6.8.3 Ligation 31
2.6.9 Transformation 32
2.6.9.1 Preparation of competent *E. coli* cells 32
2.6.9.2 Transformation of competent *E. coli* cells 32
2.6.10 Sequencing 33
2.6.11 3'-end-labelling of oligonucleotides with digoxigenin 33
2.6.12 Capillary transfer of DNA to a nylon membrane (Southern blot) 33
2.6.13 RT-PCR Southern blot 34
2.6.14 Electrophoretic Mobility Shift Assay (EMSA) 34
2.6.15 Immunological detection of DIG-labelled DNA probes 35
2.6.16 *In silico* analysis of nucleotide and amino acid sequences 35
2.7 Biochemical methods 37
2.7.1 Heterologous expression of *T. tenax* proteins in *E. coli* 37
2.7.2 Preparation, enrichment and purification of the recombinant enzymes 37

2.7.2.1 Enrichment of the recombinant Csa5 protein 37
2.7.2.2 Enrichment of the recombinant Csa2 protein 38
2.7.2.3 Enrichment of the recombinant Csa3 protein 38
2.7.2.4 Isolation and solubilisation of proteins from inclusion bodies 40
2.7.2.5 Reconstitution of protein complex CasA1 40
2.7.2.6 Reconstitution of protein complex CasA2 41
2.7.2.7 Purification of His-tagged recombinant enzymes 42
2.7.3 Protein quantitation 42
2.7.4 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) 43
2.7.4.1 SDS-PAGE assembling and electrophoresis 43
2.7.4.2 Coomassie staining 43
2.7.4.3 Molecular mass determination of proteins under denaturing conditions 43
2.7.5 Molecular mass determination of proteins under native conditions 44
2.7.6 Determination of the amino acid sequence 44

3 RESULTS 45
3.1 CRISPR loci in *T. tenax* 45
3.1.1 Identification and features of *T. tenax* CRISPR 45
3.1.2 Assignment of all spacers of the seven CRISPR loci 47
3.1.3 Analyses of small RNAs in *T. tenax* by Northern blots 49
3.2 Identification and transcription of *T. tenax cas* genes 54
3.2.1 Identification of *cas* genes in the *T. tenax* genome 54
3.2.2 Transcription of *cas* genes in dependence of abiotic stress factors 57
3.2.2.1 Cell cultivation and RNA preparation 57
3.2.2.2 Transcription under temperature and UV-light stress (RT-PCR Southern blot I) 59
3.2.2.3 Transcription under ionic stress (RT-PCR Southern blot II) 62
3.2.3 Detection of polycistronic transcripts of core *cas* genes 65
3.3 Functional analysis of the *T. tenax* DNA-binding protein Csa3 68
3.3.1 Genomic context and sequence information of *csa3* 68
3.3.2 Cloning, heterologous expression and purification of *csa3* in *E. coli* 70
3.3.3 DNA-binding studies using EMSAs 72
3.4 Analyses of the archaeal CRISPR-associated complex 1 (CasA1) 74
3.4.1 Genomic context and sequence information of *casa1* 74
3.4.2 Cloning and heterologous expression of the genes of *casa1* in *E. coli* 74
3.4.2.1 Cloning and heterologous expression of Cas4 and Csa1 74
3.4.2.2 Cloning and heterologous expression of Cas1/2 and CasA1 77

3.4.3 Refolding of CasA1 complex 79
3.4.4 Functional analyses of CasA1 complex 80
3.5 Analyses of the archaeal CRISPR-associated complex 2 (CasA2) 81
3.5.1 Genomic context and sequence information of *casa2* 81
3.5.2 Cloning and heterologous expression of genes assembled in *casa2* 82
3.5.3 Refolding of CasA2 complex 85
3.5.4 Functional analyses of CasA2 complex 88

4 DISCUSSION 91
4.1 CRISPR loci 91
4.1.1 Features of CRISPR in *Archaea* 91
4.1.2 Conservation of leader sequences in *T. tenax* 93
4.1.3 Origin of spacers 94
4.1.4 Transcription of CRISPR loci 95
4.2 *cas* genes and their transcription 99
4.2.1 Genomic context of *cas* genes in *Archaea* 99
4.2.2 Transcription of *cas* genes 102
4.2.3 Is Csa3 a transcriptional regulator for *cas* genes? 105
4.3 The CRISPR/Cas system in prokaryotes 107
4.3.1 The different roles of CRISPR/Cas systems 108
4.3.2 Function of Cas proteins 111
4.3.3 The CRISPR/Cas system in *T. tenax* – A working model 115

5 SUMMARY 118

6 LITERATURE 121

ABBREVIATIONS 142

APPENDIX 146

LIST OF FIGURES

Fig. 1.1 Phylogenetic tree of the three domains of life 1
Fig. 1.2 Electron micrograph of *Thermoproteus tenax* Kra1 6

Fig. 2.1 Blot assembly for capillary transfer of RNA from an agarose gel to a positively-charged membrane (Northern blot) 19
Fig. 2.2 Reaction scheme for the overlap extension PCR 31

Fig. 3.1 Distribution of the length of spacer sequences in the seven CRISPR loci of *T. tenax* 46
Fig. 3.2 Numbering of spacers of the seven CRISPR loci 47
Fig. 3.3 Methylene blue stained Northern blots of separated small and large RNA. 50
Fig. 3.4 Northern blot analyses of small RNA from the seven CRISPR loci of *T. tenax* 52
Fig. 3.5 Secondary structure prediction of repeat sequences 53
Fig. 3.6 Genome organisation of *cas* genes and CRISPR loci in *T. tenax* 55
Fig. 3.7 Standard growth curve of *T. tenax* 58
Fig. 3.8 Qualitative control of total RNA from control and stressed cultures 59
Fig. 3.9 RT-PCR Southern blots I and data interpretation of signals 62
Fig. 3.10 Batch cultures of *T. tenax* with and without NaCl 63
Fig. 3.11 RT-PCR Southern blots II and data interpretation of signals 65
Fig. 3.12 Genome organisation and detection of polycistronic mRNAs *casa1* and *casa2* 67
Fig. 3.13 Putative BRE-site and TATA-box elements of promoter regions *casa1*, *casa2* and *csa3* 69
Fig. 3.14 Agarose gel electrophoresis of Csa3 71
Fig. 3.15 Purification of the recombinant Csa3 71
Fig. 3.16 EMSA with Csa3 and the promoter spanning DNA regions *casa1* and *casa2* 72
Fig. 3.17 EMSA with Csa3 using *casa1* and *casa2* promoter region in presence of Poly dI-dC as competitor DNA 73
Fig. 3.18 Purification of the recombinant Cas4 76
Fig. 3.19 Expression of Cas1/2: Influence of the culture medium 78
Fig. 3.20 *In vitro* reconstitution of the CasA1 complex 80
Fig. 3.21 Test for RNA nuclease activity of CasA1 81
Fig. 3.22 Purification of the recombinant Cas3 protein 84
Fig. 3.23 SDS gel electropherogram documenting the purification of the recombinant Csa5 and Csa2 proteins from *T. tenax* 85

Fig. 3.24 *In vitro* reconstitution of the CasA2 complex....87
Fig. 3.25 SDS gel electropherogram of the refolded CasA2 complex separated by gel filtration on superose™ 88
Fig. 3.26 Assay for RNA nuclease activity of CasA2 with long and short RNA....89
Fig. 3.27 EMSA with CasA2 and short CRISPR RNA (repeat + spacer 4.22) 90

Fig. 4.1 Correlation of number of CRISPR or spacers and optimal growth temperatures in 24 archaeal organisms....93
Fig. 4.2 Multiple sequence alignments of leader sequences 94
Fig. 4.3 Hypothesis for generating small CRISPR RNAs in *T. tenax* 97
Fig. 4.4 Conservation of crenarchaeal repeat sequences....98
Fig. 4.5 Genomic context analyses of *cas* genes in prokaryotes 100
Fig. 4.6 Correlation of *cas* genes and optimal growth temperatures in 24 archaeal organisms 102
Fig. 4.7 Multiple sequence alignments of archaeal Csa3 homologs 106
Fig. 4.8 Multiple sequence alignment of Cas4 112
Fig. 4.9 Comparison of *T. tenax* Cas3 to DEAD/DEAH box helicase motifs 113
Fig. 4.10 Gene organisation of *casa2* in different archaeal organisms 114
Fig. 4.11 Working model for the function of the CRISPR/Cas system in *T. tenax* ...117

LIST OF TABLES

Tab. 2.1 Sources of the used plasmids....14
Tab. 2.2 Constructed recombinant vectors 14
Tab. 2.3 Primer sets of template generation for *in vitro* transcription....20
Tab. 2.4 Standard PCR assay 29
Tab. 2.5 Vector primer for DNA sequencing reactions....33
Tab. 2.6 Stepwise dialysis protocol for refolding of complex CasA2....42

Tab. 3.1 The seven CRISPR loci of *T. tenax* 46
Tab. 3.2 Preferred sequence similarities of *T. tenax* CRISPR spacers 48
Tab. 3.3 Quantitative analysis of RNA preparations isolated with *mir*Vana™ miRNA Isolation Kit 49
Tab. 3.4 Antisense spacer probes for Northern blot analyses of the seven CRISPR loci....51
Tab. 3.5 Gene specific primers used in secondary cDNA generation....60
Tab. 3.6 PCR primers for detection of polycistronic mRNAs *casa1* and *casa2* 66

Tab. 3.7 Oligonucleotides for amplification of the promoter spanning regions *casa1* and *casa2*.....69
Tab. 3.8 Primer sets for amplification of *csa3*.....70
Tab. 3.9 Primer sets for amplification of *cas4* and *csa1* 75
Tab. 3.10 PCR settings for amplification of *cas4* and *csa1*.....75
Tab. 3.11 Expression conditions modified for recombinant expression of Cas4.....76
Tab. 3.12 Primer sets for amplification of *cas1*/*2* and operon *casa1* 77
Tab. 3.13 Primer sets for amplification of *casa2* genes 83
Tab. 3.14 PCR settings for amplification of *casa2* genes 83

Tab. 4.1 Comparison of 24 archaeal CRISPR arrays 92
Tab. 4.2 Comparison of 24 archaeal organisms and their *cas* genes.....101

1. INTRODUCTION

Archaea were first classified as a separate domain of prokaryotes by Carl Woese and George E. Fox (Woese and Fox, 1977; Woese *et al.*, 1990) based on 16S rRNA sequence analyses. Today, all living organisms are allocated to one of the three major lines of life, which descendents form the domains of *Eucarya*, *Bacteria* and *Archaea* (Fig. 1.1). *Archaea* are subdivided into the phyla *Euryarchaeota* and *Crenarchaeota*. Recently, a further, deeply rooting branch (the branch of "Nanoarchaea") has been discovered (Huber *et al.*, 2002).

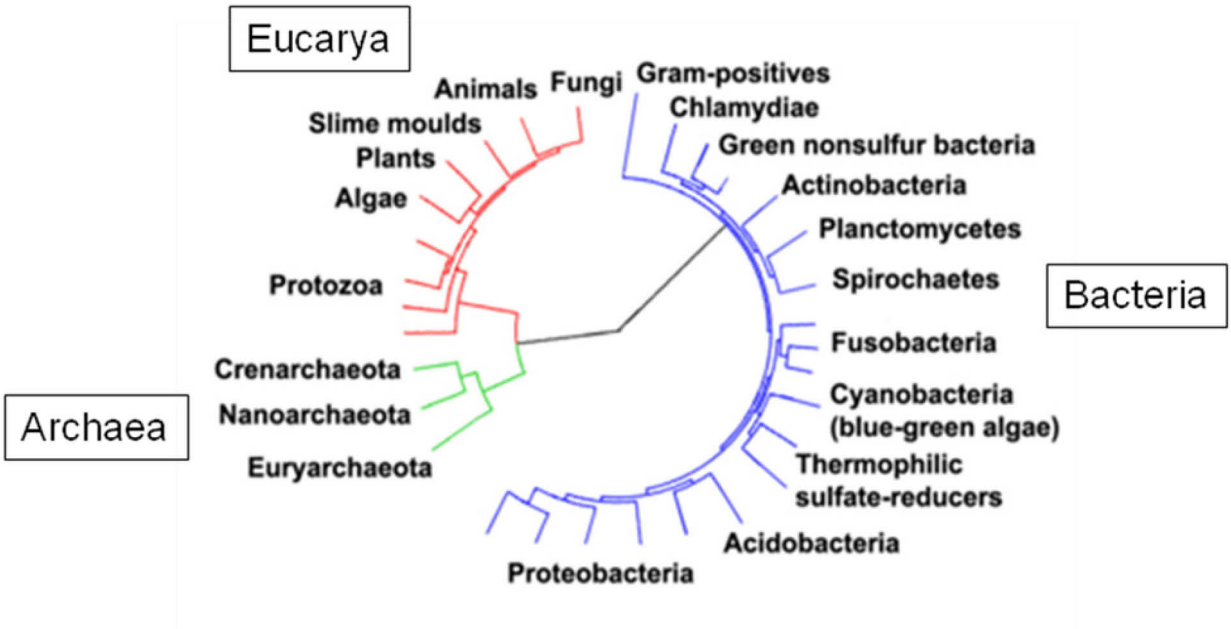


Fig. 1.1: Phylogenetic tree of the three domains of life. Tree based on 16S rRNA sequence analyses (Ciccarelli *et al.*, 2006, modified).

Archaea thrive in a broad range of habitats and may contribute to about 20 % of the total biomass on earth (De Long and Pace, 2001). Since the first archaeal organisms were isolated from extreme environments, it has been suggested that the extremophilic lifestyle is generally representative for that group of organisms (Valentine, 2007). Some species thrive at high temperatures over 100°C or cold temperatures down to 0°C, in extremely alkaline or acidic waters (pH 1-2), or saline environments (>30 % (w/v) sodium chloride). Respective organisms were isolated from the deep sea (black smoker), geysers, salt lakes, hot acidic springs and sulphuric waters. But in the meantime archaeal members could also be isolated from rather "normal", i.e. less spectacular habitats such as digestive tracts of cows or termites, anaerobic mud, oceans, soils and marshlands demonstrating their exceptional adaptability of these organisms. Thus, *Archaea* represents attractive objects for basic research (concern-

ing biochemical, molecular biological or ecological and evolutionary aspects), as well as for applied research (concerning biotechnological applications of their resistant physiology and macromolecular equipment). Although some archaeal members have been studied in detail, the knowledge about the molecular background of the archaeal operational and information processing abilities is still rather scarce as compared to members of *Bacteria* and *Eucarya*. The phylum of the *Crenarchaeota* mainly includes hyperthermophilic and thermoacidophilic organisms, whereas the members of *Euryarchaeota* represent a more diverse group including methanogenic, halophilic and hyperthermophilic species.
Present data deduced from comparisons of complete genome sequences indicate that archaeal organisms represent a mosaic of bacterial and eukaryotic features, as their mechanisms of information processing are more related to *Eucarya*, whereas their metabolic functions resemble those of *Bacteria* (Koonin and Galperin, 2003; Makarova and Koonin, 2003; Allers and Mevarech, 2005).

The basal archaeal transcription apparatus is more closely related to the core components of *Eucarya* than of *Bacteria* (Bell and Jackson, 1998; 2001), characterised by a type II-like DNA dependent RNA polymerase (RNAP), which resembles the eucaryal nuclear RNAPs in subunit composition (Zillig *et al.*, 1979) and homologs of several basic eucaryal transcription factors, including the TATA-box binding protein (TBP) and transcription initiation factor IIB (TFIIB; Bell and Jackson, 1998; Bell *et al.*, 2001). On the other hand, the majority of genes in *Archaea* (50-70 %) are organized in operon structures (similar to *Bacteria*) containing on average three genes and subsequently producing polycistronic mRNA transcripts (Torarinsson *et al.*, 2005). The main sequence logos of the promoter are the box A motif (TATA-box) and the TFIIB recognition element (BRE-site), interacting with TFB (Bell *et al.*, 1999). Surprisingly, most of the so far identified transcriptional regulators are bacterial-like (Kyrpides and Ouzounis, 1999; Aravind and Koonin, 1999). Therefore, transcriptional regulation in *Archaea* is accomplished by bacterial-type regulators that interact with the eucaryal-like transcription machinery. Several transcriptional regulators have been characterised in *Archaea*, e.g. the metal dependent repressor 1 (MDR1) from *Archaeoglobus fulgidus*, a homolog of metal-dependent bacterial repressors (Bell *et al.*, 1999). A maltose-specific regulator (TrmB) for the trehalose/maltose transport

operon from *Thermococcus litoralis* has also been characterized, inhibiting transcription counteracted by maltose and trehalose (Lee *et al.*, 2003).

Repeats in prokaryotic genomes

Genomes are more than the sum of an organism's genes and have traits that can be studied such as chromosome number, genome size, gene order, codon usage bias and GC-content. To determine what mechanisms could have produced the great variety of genomes existing today, the comparison of genome features and structures is necessary (Gregory, 2004).

The amplification of genetic material leads to duplicated sequences resulting in an extension of short repeats, duplication of gene cluster or entire chromosomes and even entire genomes (Treangen *et al.*, 2009). Such duplications provide the basic for functional novelties in the genomes of animals and plants. In prokaryotes, the amplifications of genetic material compete with horizontal gene transfer for the major role among the mechanisms allowing the acquisition of novel functions. In comparison to *Eucarya*, *Bacteria* and *Archaea* have small compact genomes with typically 85-90 % of sequence coding for proteins or stable RNA molecules. Most of the remaining genome corresponds to regulatory regions. Although prokaryotes do not achieve the number of repeats found in eukaryotes, their genomes often contain hundreds of large repeats. The proliferation of selfish elements, such as retrotransposons or insertion sequences (ISs) in genomes is one of the main sources of repeats. ISs constitute the most frequent and repetitive transposable elements in prokaryotic genomes and are the simplest autonomous mobile genetic elements ranging from 0.7 to 3.5 kbp (Mahillon and Chandler, 1998). Interspersed repeats are another family of repeat sequences. They are widespread in bacterial and archaeal chromosomes and plasmids, typically located in intergenic regions (Ogata *et al.*, 2000). Examples for that family are repetitive extragenic palindromic units (REP), bacterial interspersed mosaic elements (BIME) or clustered regularly interspaced short palindromic repeats (CRISPR). This work gives attention to the characterization and function of this novel family of repeat elements in prokaryotes.

The CRISPR/Cas system in prokaryotes

CRISPR sequences are composed of direct repeats, repeated up to 250 times, ranging in size from 24 to 47 nt that are separated by similarly sized nonrepetitive spacers

(Jansen *et al.*, 2002). While the direct repeats are similar within a single locus, they are different between CRISPR loci of the same genome. In the majority of cases, the direct repeats are highly conserved while the spacers are very diverse within a given locus, even among strains of the same species (Pourcel *et al.*, 2005). CRISPR sequences are widely distributed in *Bacteria* and *Archaea*: One or more loci are found in 40 % of the bacterial genomes sequenced so far and in nearly all *Archaea* (Grissa *et al.*, 2007). CRISPR arrays can rapidly evolve and are often highly different between otherwise closely related strains. Similarity searches of spacer sequences of CRISPR showed that some spacer match to phages and other extrachromosomal elements, such as plasmids or transposons, but also to chromosomal DNA (Mojica *et al.*, 2005; Bolotin *et al.*, 2005). As another characteristic feature of CRISPR, an AT-rich sequence of up to 550 bp is located at one flanking site of a CRISPR locus and denoted as the leader sequence. Similar to repeats, these sequences lack an open reading frame and are conserved in the same species.

Originally, four genes, always located near to a CRISPR locus and only found in species containing CRISPR sequences, were identified in numerous prokaryotic genomes and therefore designated as CRISPR-associated (*cas*) genes (Jansen *et al.*, 2002). Meanwhile, in addition to the four designated *cas* genes, further genes assembled to 41 *cas* gene families were identified and described (Haft *et al.*, 2005). On the basis of genomic context comparisons, eight subtypes were defined in prokaryotes with clearly different sets of *cas* genes. Each subtype contains at least two of the core genes *cas1-6* and up to eight different subtype-specific *cas* genes. The respective subtypes were named after a well-known representative, e.g. subtype-specific *cas* genes of the subtype *Aeropyrum pernix* are termed *csa* or of subtype *Mycobacterium tuberculosis* are termed *csm*. Generally, these *cas* genes are assumed to be involved in the function and maintenance of CRISPR. Several hypotheses about the function of the CRISPR/Cas system have been proposed. Based on the observation that an increase in the copy number of the repeats results in altered replicon segregation in *Haloferax volcanii*, an involvement of CRISPR sequences in replicon partitioning was suggested (Mojica *et al.*, 1995). As many *cas* genes showed similarity to DNA-modifying enzymes, such as nucleases or helicases, it has been suggested that the CRISPR/Cas system is functioning as a novel DNA repair system (Makarova *et al.*, 2002). Recently, it was demonstrated that, in response to phage infection, *Bacteria* integrate new spacers into their CRISPR

arrays, which results in CRISPR-mediated phage resistance. The new integrated spacers were derived from the genome of the challenging phage, due to 100 % identity of spacers and phage sequences (Barrangou *et al.*, 2007). The inheritable nature of CRISPR arrays and the rapidly evolving structures, declare them as perfect targets for evolutionary strain typing (spacer-oligotyping or “spoligotyping”) and comparative genomic studies, especially for pathogenic microorganisms (Crawford, 2003; Brudey *et al.*, 2006). Further potential future applications are the engineered defence against viruses or the selective silencing of endogenous genes (Sorek *et al.*, 2008).

However, the role of CRISPR arrays in microbial genomes and the mechanisms that underlie CRISPR function are mostly uncharacterized, especially in *Archaea*. Also, whether different CRISPR systems contain different functionalities is still puzzling (Sorek *et al.*, 2008). The main difference of archaeal repeat cluster in comparison to *Bacteria* are the tendency to have more and larger CRISPR loci and *cas* genes. Up to 1 % of the chromosome can be covered by CRISPR in thermophilic archaeal organisms (Nelson *et al.*, 1999; Lillestol *et al.*, 2006).

Strong evidence was provided that CRISPR loci are transcribed and processed into a series of smaller RNAs, corresponding in length to multiples of repeat-spacer units (Tang *et al.*, 2002; 2005). The transcription start site is located immediately upstream of the first repeat, preceded by archaeal BRE/TATA motifs within the leader sequence (Lillestol *et al.*, 2009). In *Archaeoglobus fulgidus* Northern blot analyses with probes targeting the CRISPR repeat sequence (30 nt) revealed sizes of 272, 204, 136 and 68 nt (Tang *et al.*, 2002). Similar results were obtained in *S. solfataricus* with a 24 nt long repeat probe (540, 360, 180 and 60 nt; Tang *et al.*, 2005). Based on the increasing numbers of available prokaryotic genome sequences, an ever-increasing number of regulatory RNAs is identified, including mRNA leaders that affect expression in *cis*, small RNAs that bind to proteins or complementary target RNAs and small CRISPR RNAs. In general, regulatory RNAs can modulate transcription, translation, mRNA stability, and DNA maintenance or silencing. They achieve these effects by various mechanisms, including changes in RNA conformation, protein binding, base pairing with other RNAs, and interactions with DNA (Waters and Storz, 2009). One of the simple RNA regulatory elements are riboswitches located within mRNA transcripts that can adopt different conformations in response to sig-

nals, such as temperatures or presence/absence of small molecule ligands (Grundy and Henkin, 2006). Most characterised are small RNAs, which regulate gene expression by base pairing with mRNAs. *Cis*-encoded small RNAs are located on the complementary DNA strand of the target mRNA and resides often on plasmids or other mobile genetic elements (Brantl, 2007). Another class of base pairing RNAs are the *trans*-encoded small RNAs, which share only limited complementarities with their target mRNAs. These small RNAs regulate the translation and/or stability of target mRNAs and are, in many respects, functionally analogous to eukaryotic miRNAs (Aiba, 2007).

Thermoproteus tenax

T. tenax strain Kra1 was the first described hyperthermophilic Archaeum belonging to the phylum of the *Crenarchaeota* and was isolated from a solfatare in Iceland (Zillig *et al.,* 1981; Fischer *et al.*, 1983). *T. tenax* is a rod-shaped, strictly anaerobic, sulphur reducing organism with optimal growth at 86°C and pH 5.5. The organism is able to grow chemolithoautotrophically on carbon dioxide and hydrogen as well as chemoorganoheterotrophically in the presence of various organic substrates, e.g. glucose, starch, amylase or glycerol (Zillig *et al.*, 1981; Fischer *et al.*, 1983).

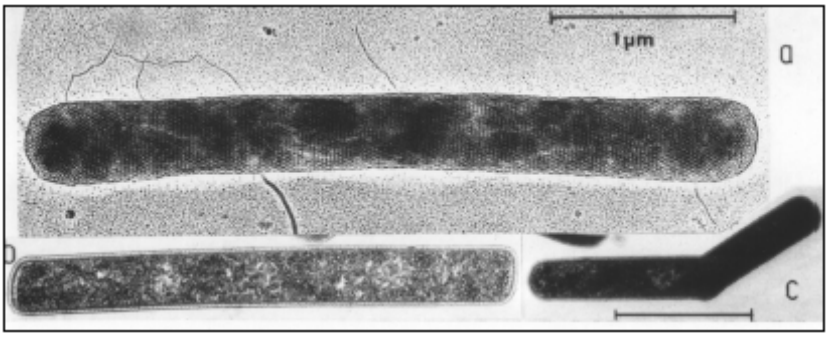


Fig. 1.2: Electron micrograph of *Thermoproteus tenax* Kra1. (Zillig *et al.*, 1981)

Several previous studies focused on mechanism and regulation of the central metabolic pathways of this organism (Siebers and Hensel, 1993; Siebers *et al.*, 1997, 2004; Brunner *et al.*, 1998, 2001; Schramm *et al.*, 2000; Ahmed *et al.*, 2004, 2005; Tjaden *et al.*, 2006). *T. tenax* uses modified versions of two classical pathways known from *Bacteria* and *Eucarya* for the metabolism of carbohydrates: The reversible Embden-Meyerhof-Parnas (EMP) pathway and the branched Entner-Doudoroff

(ED) pathway. Interestingly, this organism follows a different strategy to adapt the metabolic activity to changed trophic conditions. Instead of the common regulation sites of the reversible EMP pathway at the level of glucose / glucose 6-phosphate and fructose 6-phosphate / fructose 1,6-bisphosphate interconversion, *T. tenax* possess control points at the level of glyceraldehyde 3-phosphate / 3-phosphoglycerate interconversion (by three glyceraldehyde 3-phosphate converting enzymes: i) the allosteric non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase, GAPN; ii) the classical phosphorylating glyceraldehyde 3-phosphate dehydrogenase, GAPDH; iii) the non-phosphorylating ferredoxin-dependent glyceraldehyde 3-phosphate oxidoreductase, GAPOR) and at the level of the phosphoenolpyruvate / pyruvate conversion (by: i) the catabolic pyruvate kinase, PK; ii) the reversible pyruvate phosphate dikinase, PPDK, which catalyses the interconversion of phosphoenolpyruvate and pyruvate, but mainly functioning as a catabolic enzyme; iii) the anabolic phosphoenolpyruvate synthetase, PEPS). Both control points are equipped with three enzymes and allow short-term and long-term adaptation by regulation on protein and gene level.

The *T. tenax* genome sequence was deciphered in collaboration with Dr. H-P Klenk (e.gene, Feldafing, Germany) and Prof. Dr. SC Schuster (MPI für Entwicklungsbiologie, Tübingen, Germany / Pennsylvania State University, USA). This sequence has been used for the identification of CRISPR loci and *cas* genes of *T. tenax* (Siebers *et al*., 2010; submitted).

Aims of the work

The role of CRISPR in microbial genomes and the mechanisms that underlie CRISPR function are largely uncharacterised, especially in *Archaea*. For a detailled analysis of the archaeal CRISPR system the well studied hyperthermophilic Archaeum *Thermoproteus tenax* has been chosen.

These studies focus on the identification of basic features of the CRISPR/Cas system in *T. tenax* such as CRISPR loci and gene organisation of *cas* genes, as well as on analyses of CRISPR-derived small RNAs to answer basic question about the function of the CRISPR/Cas system in *Archaea*.

The studies should also include the functional characterisation of the involved Cas proteins to provide insights into the molecular mechanisms of the CRISPR/Cas system in *T. tenax.*

Further experiments of the transcript levels of *cas* genes under different growth conditions should give indications about the sensitivity of the CRISPR/Cas system towards abiotic signals and about possible response regulators.

2. MATERIALS AND METHODS

2.1 Chemicals, enzymes, kits and consumables

Acrylamide, N,N-methylenebisacrylamide	Roth GmbH, Karlsruhe
Alkaline phosphatase, calf intestinal (HC)	Fermentas GmbH, St. Leon-Rot
Anti-digoxigenin-AP, Fab-fragments	Roche Diagnostics GmbH, Mannheim
Antibiotics	Sigma-Aldrich, Taufkirchen
Blocking Reagent	Roche Diagnostics GmbH, Mannheim
Bradford Reagent	BioRad, München
CDP-Star (Tropix)	Applied Biosystems, Darmstadt
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Comassie-Brilliant-Blue R and G	SERVA GmbH, Heidelberg
Dialysis tubes	Roth GmbH, Karlsruhe; Serva GmbH, Heidelberg
Diethyl pyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
DIG Oligonucleotide 3'-end Labelling Kit	Roche Diagnostics GmbH, Mannheim
DIG RNA-Labelling-Kit (SP6/T7)	Roche Diagnostics GmbH, Mannheim
DNA Ladder Mix, GeneRuler™	Fermentas GmbH, St. Leon-Rot
DNA Ladder Ultra Low Range, GeneRuler™	Fermentas GmbH, St. Leon-Rot
DNA Molecular Weight Marker II, DIG	Roche Diagnostics GmbH, Mannheim
DNA-Polymerase, *Pyrococcus furiosus*	Fermentas GmbH, St. Leon-Rot
DNA-Polymerase, *Thermus aquaticus*	Fermentas GmbH, St. Leon-Rot
DNase, on-column	QIAGEN, Hilden
dNTP Mix	Fermentas GmbH, St. Leon-Rot
DNAzol Reagent	Invitrogen GmbH, Karlsruhe
Isopropyl-ß-thiogalactopyranosid (IPTG)	AppliChem GmbH, Darmstadt
*mir*vana™ miRNA Isolation Kit	Applied Biosystems, Darmstadt
Ni-NTA agarose	QIAGEN, Hilden
Oligonucleotides	Sigma-Aldrich, Taufkirchen
Protein Ladder, 10-200 kDa	Fermentas GmbH, St. Leon-Rot
Roti®-Nylon plus, nylon membrane	Roth GmbH, Karlsruhe
QIAGEN Plasmid-Kit, QIAquick Gel Extraction Kit, QIAquick Nucleotide Removal Kit, QIAquick PCR Purification Kit	QIAGEN, Hilden

Restriction endonucleases	Fermentas GmbH, St. Leon-Rot
Reverse transcriptase	Fermentas GmbH, St. Leon-Rot
RNA ladder, high range	Fermentas GmbH, St. Leon-Rot
RNase-ExitusPlus™	AppliChem GmbH, Darmstadt
RNeasy Mini Kit	QIAGEN, Hilden
SDS	SERVA GmbH, Heidelberg
T4-DNA-Ligase	Fermentas GmbH, St. Leon-Rot
T7-RNA-Polymerase	Fermentas GmbH, St. Leon-Rot
T7 *in vitro* transcription Kit	Fermentas GmbH, St. Leon-Rot
TRIzol Reagent	Life Technologies, Karlsruhe
Whatman GB 004, 3MM	Schleicher & Schuell GmbH, Dassel

Remaining chemicals unlisted above were purchased from Gerbu (Gaiberg), Sigma-Aldrich (Taufkirchen), Roth GmbH (Karlsruhe), Difco Laboratories (Augsburg) and VWR International (Darmstadt) in analytical grade as well as organic solvents purchased from J.T. Baker B.V. Deventer (NL).

2.2 Instruments

Aqua bidest. water system	Milli Q Biocel A10, Millipore, Schwalbach
Agarose gel electrophoreses	Agagel Mini, Biometra, Göttingen; Power supply: Consort E143 (MS Laborgeräte)
Autoclave	Zirbus LVSA 40/60, Bad Grund, H+P Varioklav 75 T, Oberschleißheim
Benchtop heater	BT3, Grant Instruments, Cambridge, UK
Cell disruption/homogenisation	French Press, SLM Aminco Instruments Inc., distributed by Sopra GmbH, Büttelborn; hand held glass-teflon homogeniser, B. Braun AG, Melsungen
Cell counting	Assistent Neubauer chamber, LO-Laboroptik GmbH, Friedrichsdorf
Centrifuges	Bench centrifuges: Hettich Universal centrifuge 32R; Biofuge® pico Biofuge A, Heraeus Instruments, Eppendorf 5412 and Eppendorf 5804 R, Hamburg; large centrifuge: Mega-

	fuge 1.0 R, Kendro, Langenselbold; Avanti J-25, Beckmann, München; Ultracentrifuge: L8-80, Beckman Coulter GmbH, Krefeld
Chemiluminescence detector and gel documentation	Fusion FX 7, Vilber Lourmat, Eberhardzell; ChemiDoc Gel Documentation System, Bio-Rad Laboratories GmbH, München; P91W B/W thermal printer, Mitsubishi
Chromatography Columns and Media	GE Healthcare Europe GmbH, Freiburg
Gas exchange system	Self-building and Air Liquide, Düsseldorf
High pressure liquid chromatography (HPLC)	HPLC System Gold, 126NM Solvent Module, 166NM Detector, Beckman Coulter, Krefeld
Hybridisation oven	OV3 Mini hybridisation oven and Compact Line OV4, Biometra, Göttingen
Incubators	Certomat H / Certomat R, B. Braun AG, Melsungen; Minitron Incubator Infors AG, Bottmingen, Basel, Switzerland
Microscopes	Olympus BH-2 RFCA and Olympus CHT, Olympus, Hamburg
Photometer	Specord 200, Analytik Jena AG, Jena with WinASPECT Spectralanalysis-Software; BioPhotometer Eppendorf, Hamburg
Polyacrylamide gel electrophoresis	Minigel-Twin, Biometra, Göttingen; Power supply: Consort E835, MS Laborgeräte
Temperature probe	P510 Temp, Dorstmann Electronic GmbH, Wertheim
Thermal cycler	MJ Mini and iCycler, BioRad Laboratories GmbH, München
UV light	UVVIS 254nm, Sarstedt, Nürnbrecht
Vacuum centrifuge	SpeedVac Concentrator, Savant Farmingdale, UK
VIVASPIN columns	VIVASCIENCE, Sartorius group, Stonehouse, UK

2.3 Strains and culture conditions

Thermoproteus tenax Kra1 strain; DSMZ 2078 (Zillig *et al*., 1981)
Escherichia coli K12 DH5α strain; DSMZ 6897 (Hanahan, 1983)
Escherichia coli BL21 (DE3); Novagen (Studier & Moffatt, 1986)
Escherichia coli Rosetta (DE3); Novagen
Escherichia coli BL21 (DE3) pLysS; Novagen (Studier & Moffatt, 1986)
Escherichia coli BL21-CodonPlus (DE3) pRIL; Stratagene (Carstens & Waesche, 1999)

2.3.1 Culture conditions for growing *T. tenax*

Cultures of *Thermoproteus tenax* Kra1 were grown heterotrophically at 86°C and pH 5.5 in a complex medium (modified to Brock *et al*., 1972) containing (per litre):
1.3 g $(NH_4)SO_4$, 280 mg KH_2PO_4, 250 mg $MgSO_4$ x $7H_2O$, 70 mg $CaCl_2$ x $2H_2O$, 20 mg $FeSO_4$ x $7H_2O$, 1.8 mg $MnCl_2$ x $4H_2O$, 4.5 mg $Na_2B_4O_7$ x $10H_2O$, 0.22 mg $ZnSO_4$ x $7H_2O$, 0.05 mg $CuCl_2$, 0.05 mg Na_2MoO_4 x $2H_2O$, 0.03 mg $VOSO_4$ x H_2O, 0.01 mg $CoSO_4$ x $7H_2O$, 1 mg resazurin. Afterwards the medium was heated up to 86°C and adjusted at pH 5.5 with KOH. To achieve an anaerobic atmosphere, 20 or 500 ml medium was filled in 100 ml or 2 l bottles respectively, closed gastight with butyl plugs and purged with pure N_2. Subsequently, the medium was autoclaved and stored at room temperature.
Before use, 1 g/l dispersed elemental sulphur (S^0) was added. 0.2 g/l yeast extract and 0.3 g/l L-cysteine were passed through a sterile filter and injected. The L-cysteine acted as a reducing agent and guaranteed anaerobic growth conditions. After a preheating time of 2 h the medium was inoculated with 4 % of a primary culture (1 x 10^8 cells x ml^{-1}) and incubated at 86°C without agitation. Growth of the culture was monitored by counting the cells with a Neubauer chamber. After 3-4 days, the growth of the heterotrophic cultures was stopped in late exponential phase (8-9 x 10^7 cells x ml^{-1}) by quickly cooling down to 4°C and centrifugation (10,000 x g, 20 min, 4°C). Finally the cells were pooled and stored at -80°C.

2.3.1.1 Culture conditions at increased ionic strength

Growth of *T. tenax* is inhibited by higher ionic strength in the medium, 500 mM NaCl leads to dieback of the culture. To induce stress conditions, NaCl concentration up to

50-150 mM was increased by adding respective volumes of a sterile 5 M NaCl stock solution to an exponentially grown *T. tenax* culture with cell counts of 8-9 x 10^7 cells x ml^{-1}. Subsequently, the medium was incubated at 86°C for 3-6 h without agitation. The growth of the culture was stopped by cooling down to 4°C and cells were harvested by centrifugation (s. 2.3.1).

2.3.1.2 Culture conditions under UV-radiation

The *T. tenax* cultures were grown to 8-9 x 10^7 cells x ml^{-1}, cooled down and centrifuged (10,000 x *g*, 20 min, 4°C). Afterwards the pellet was resuspended in 50 ml medium, transferred to an anaerobic chamber and poured in a Petri dish (12 x 12 cm) with a level of 3 mm. The cells were irradiated at RT for 30 sec to 2 min at α = 254 nm with a portable UV-lamp (5 J/m^2 or 20 J/m^2, respectively), while swaying the culture carefully. The cells were reinjected into 500 ml of fresh and pre-heated (86°C) medium with a syringe, incubated for 3 h at 86°C without agitation and harvested as described above (s. 2.3.1).

2.3.1.3 Culture conditions at sub- and supraoptimal temperatures

The *T. tenax* cultures were grown in an incubator at 86°C to 8-9 x 10^7 cells x ml^{-1}. The temperature variation was achieved by a followed incubation in a water bath at 70°C or 91°C, respectively. Adjustment of the temperature was carried out for 30 min, in which a reference vessel filled with water served as a temperature control for the culture. After the temperature adjustment, the cultures were further incubated for 3 h and harvested as described above (s. 2.3.1).

2.3.2 Culture conditions for *E. coli*

The aerobic cultivation of *E. coli* batch cultures (3-300 ml) was carried out in Erlenmeyer flasks by shaking in a rotatory shaker (200 rpm) at 37°C in lysogeny broth (LB) medium (1 % peptone, 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2) or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar). Alternative media, were 2x YT (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl (w/v), pH 7.2) and NZA (1 % N-Z-Amine A®, 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2). The aeration of mass cultures (volumes up to 6 l) was achieved by gassing with air (flow rate 50 l/min). Liquid LB medium containing the appropriate plasmid-encoded antibiotics (chloramphenicol 34 µg/ml, kanamycin 50 µg/ml and ampicillin 100 µg/ml) was inoculated with a pre-

culture (2 % (v/v)) and growth was monitored at 600 nm. Protein expression was induced at OD_{600} = 0.6-0.8 by the addition of 1 mM IPTG and incubation was continued for 3-4 hours at 37°C. The cells were then chilled on ice, harvested by centrifugation (6,000 x *g*, 15 min, 4°C) and stored at -80°C. The *E. coli* strain K12 DH5α was used for cloning, storage and preparation of plasmid DNA. The pET Vector System (Novagen) and the strains *E. coli* BL21(DE3), BL21(DE3) pLysS, BL21-CodonPlus(DE3) pRIL and Rosetta(DE3) were used for the heterologous expression of recombinant *T. tenax* proteins.

2.4 Plasmids and constructed recombinant vectors

Tab. 2.1: Sources of the used plasmids

Vector	Resistance	Application	Source
pBluescript II KS+	Amp^r	cloning	Stratagene
pET15b	Amp^r	heterologous expression	Novagen
pET11c	Amp^r	heterologous expression	Novagen
pET24a	Kan^r	heterologous expression	Novagen
pLysS	Cam^r	expression of T7-lysozyme in *E. coli*	Novagen
RIL	Cam^r	expression of rare tRNA-genes *argU*, *ileY*, *leuW*	Stratagene
pSPT19	Amp^r	*in vitro* transcription	Roche

Tab. 2.2: Constructed recombinant vectors

Plasmid + Insert	Description
pET24a + *cas4*	*T. tenax cas4* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *cas4* his	*T. tenax cas4* gene + C-terminal His-Tag in pET24a, restriction sites *Nde*I / *Bam*HI
pBlue + *cas1*/*2*	*T. tenax cas1*/*2* gene in pBluescript II KS+, restriction site *Bam*HI
pET15b + *cas1*/*2*	*T. tenax cas1*/*2* gene in pET15b, restriction site *Nco*I / *Bam*HI
pET24a + *csa1*	*T. tenax csa1* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *csa1* his	*T. tenax csa1* gene + C-terminal His-Tag in pET24a, restriction sites *Nde*I / *Bam*HI

pET24a + *csa3*	*T. tenax csa3* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *csa5*	*T. tenax csa5* gene in pET24a, restriction sites *Nde*I / *Hind*III
pET24a + *csa2*	*T. tenax csa2* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *cas5a*	*T. tenax cas5a* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *cas3*	*T. tenax cas3* gene in pET24a, restriction sites *Nde*I / *Eco*RI
pET24a + *cas3* his	*T. tenax cas3* gene + C-terminal His-Tag in pET24a, restriction sites *Nde*I / *Eco*RI
pET24a + *cas3hd*	*T. tenax cas3hd* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *csa4*	*T. tenax csa4* gene in pET24a, restriction sites *Nde*I / *Bam*HI
pET24a + *casa1*	2.3 kb-operon of *T. tenax* genes *cas4*, *cas1*/*2* and *cas4a* in pET24a, restriction sites *Nde*I / *Bam*HI

2.5 Working with RNA

2.5.1 Treatment of solutions, glassware and equipment

For protection against RNases, all applied buffers and solutions were mixed with 0.1% (v/v) diethyl pyrocarbonate (DEPC), incubated overnight at RT and autoclaved to remove traces of DEPC. Since reagents with an amino group, such as Tris or SDS, cannot be treated directly with DEPC, these solutions were mixed with DEPC-H_2O_{bidest}, incubated overnight at RT and subsequently autoclaved. Glassware was heat sterilised by incubation at 210°C for at least 2 hours before use. Non-disposable plastic ware was treated with 3% H_2O_2 or RNase Exitus Plus™ (AppliChem). Pipette tips, as well as reaction tubes, were autoclaved twice.

2.5.2 Isolation of small and total RNA from *T. tenax*

2.5.2.1 Isopropanol precipitation

Total RNA was prepared from heterotrophically grown *T. tenax* cells harvested by using TRIzol reagent (Invitrogen), a monophasic solution of phenol and guanidinium isothiocyanate, followed by organic extraction and alcohol precipitation of the RNA as described by Chomczynski (1997).

0.1 g of *T. tenax* cells were homogenised for 3 min on ice in 1 ml TRIzol using a hand-held glass-teflon homogeniser and then incubated for 5 min at RT. After the addition of 200 µl chloroform, the reaction tube was mixed vigorously, incubated for

2–3 min at RT and centrifuged (17,500 x *g*, 15 min, 4°C). The upper, yellowish aqueous phase (approximately 500-600 µl) containing the RNA was aliquoted to 40 µl and transferred to fresh tubes. After this step, the RNA was precipitated by the addition of 1.25 vol isopropanol, incubated for 10 min at RT and centrifuged (17,500 x *g*, 10 min, 4°C). The supernatant was discarded and the RNA was washed by adding 100 µl 70% ethanol and centrifuged (17,500 x *g*, 10 min, 4°C). The RNA was resuspended and stored in 100% ethanol at -80°C. Before use, the RNA was washed with 70% ethanol, centrifuged (17,500 x *g*, 10 min, 4°C), carefully dried at 37°C and resuspended in 20 µl DEPC-H_2O_{bidest}.

2.5.2.2 RNeasy Mini Kit (QIAGEN)

The RNeasy Mini Kit was used to prepare clean and DNA-free RNA for subsequent reverse transcription. For that purpose, the instruction's manual was modified by using cell disruption with 1.5 ml TRIzol reagent and 0.15 g of *T. tenax* cells as mentioned above. After the addition of 300 µl chloroform and centrifugation, the upper, yellowish aqueous phase was mixed with 700 µl RLT buffer containing 0.1 % α-Me and 500 µl 100 % ethanol. The sample was applied on the spin column and centrifuged (10,000 x *g*, 1 min, RT). To eliminate contamination with genomic DNA, an on-column DNase treatment was performed with slight modifications. The on-column DNase treatment was performed for 15-20 min at 30°C and the subsequent wash steps were carried out according to the instruction's manual. The RNA was finally eluted from the column by applying 30 µl of RNase-free water, incubated for 1 min at RT and centrifuged (10,000 x *g*, 1 min, RT). This elution procedure was repeated and in a third step, 30 µl of eluate was applied on the column, incubated and centrifuged in order to improve RNA recovery.

2.5.2.3 *mir*Vana™ miRNA Isolation Kit (Ambion)

For preparations of small RNA species (< 200 nt) the *mir*Vana™ miRNA Isolation Kit (Ambion) was used. The instruction´s manual was modified by homogenisation of 0.1 g *T. tenax* cells in 1 ml Lysis/Binding buffer (containing guanidinium isothiocyanate) using a hand-held glass-teflon homogeniser for 3 min on ice. After the addition of 100 µl miRNA Homogenate Additive™ and incubation for 10 min on ice, the sample was phenol/chloroform extracted. To separate the total RNA into two fractions and enrich the small RNA species, the sample was brought in a first step to 25% ethanol to im-

mobilize large RNAs on the glass-fiber filter. The small RNA species were collected in the filtrate. The ethanol concentration of the filtrate was then increased to 55%, and passed through a second glass-fiber filter to immobilize small RNAs. Both RNA fractions were washed and eluted in 100 µl Elution Solution pre-heated at 95°C.

2.5.3 Quantitative and qualitative analysis of RNA

The concentration of extracted RNA was determined photometrically at α = 260 nm. The absorption of 1 corresponds to 40 µg RNA/ml for normal preparations and 33 µg RNA/ml for preparations of RNA smaller than 200 nt, respectively (Sambrook *et al*., 1989). In addition, the A_{260}/A_{280} ratio is an indication for RNA purity. Sufficiently pure RNA preparations showed a ratio higher than 1.8, whereas ratios lower than 1.8 indicate contamination with protein or phenol. The integrity of purified RNA was checked by agarose gel electrophoresis upon ethidium bromide staining (s. 2.6.5.1), as well as by Northern blotting and methylene blue staining (s. 2.5.5).

2.5.4 Gel electrophoresis of RNA

2.5.4.1 Denaturing agarose gel electrophoresis

Electrophoresis separation of RNA longer than 200 nt was achieved under denaturing conditions in agarose-MOPS/formaldehyde gels (Staynov *et al*., 1972). For 1-1.5 % (w/v) MOPS/formaldehyde gels, 1 or 1.5 g agarose was added to 73.8 ml DEPC-H_2O_{bidest} and 10 ml 10x MOPS buffer (10x: 200 mM morpholino propane sulfonic acid (MOPS), 50 mM sodium acetate, 10 mM EDTA, pH 7.0) and microwaved until the solution became clear. After cooling down to approximately 65°C, 16.2 ml of formaldehyde (37 % (v/v)) was added and the agarose-formaldehyde mixture was poured into a gel tray. A comb (10 wells) was placed into the tray slot. The RNA samples and RNA size marker were mixed with 3 vol of freshly prepared formaldehyde loading dye (250 µl deionised formamide, 83 µl 37 % (v/v) formaldehyde, 50 µl 10x MOPS buffer, 8.5 µl 2 % (w/v) bromophenol blue and 8.5 µl DEPC-H_2O_{bidest}), incubated for 10 min at 65°C and finally chilled on ice. The solidified gel was placed in the electrophoresis chamber and was not completely covered with 1x MOPS buffer, until the RNA probes were run into the gel matrix. Electrophoresis was performed at 85 V for 1-2 h.

2.5.4.2 Denaturing polyacrylamide gel electrophoresis

Electrophoretic separation of RNA smaller than 200 nt was achieved under denaturing conditions in urea-polyacrylamide gels. Here, secondary structures of RNAs are resolved by 8 M urea and separation depends mainly on the size of RNA. Special attention must be paid on cleaning the materials for gel casting. Thus, the glass plates were carefully cleaned and heat sterilised (s. 2.5.1); spacers and combs were treated with RNase Exitus Plus™.

The polyacrylamide (acrylamide / bisacrylamide, 40 %, ratio 29:1) concentration of the gel, containing 8 M urea, 90 mM Tris, 90 mM boric acid, 2 mM EDTA pH 8, 0.03 % (v/v) APS, 0.005 % (v/v) TEMED, varied between 8 and 15 % depending on the size of the applied RNA. The freshly prepared, polymerised gel was assembled in the gel chamber (Minigel-Twin-Chamber, Biometra), filled with 1x TBE (90 mM Tris, 90 mM boric acid, 2 mM EDTA pH 8) and the voltage was adjusted to 80 V for 30 min before applying the samples. The primary reason for pre-running the gel was to remove traces of ammonium peroxodisulfate (APS). The RNA samples were mixed either with 1 vol Direct Loading Dye (80 % formamide, 10 mM EDTA, 0.05 % bromophenol blue) or 1 vol Gel Loading Buffer II (Ambion, 95 % formamide, 18 mM EDTA, 0.025 % SDS, 0.025 % xylene cyanol, 0.025 % bromophenol blue), incubated for 5 min at 95°C and placed on ice. The gel slots were rinsed thoroughly with 1x TBE, right before applying the sample to guarantee the sinking of the RNA mixture into the slots. The gel run was performed at a constant voltage of 80 V for 2 h.

2.5.5 Capillary transfer of RNA to a nylon membrane (Northern blot)

The electrophoretically separated RNA was transferred from the agarose-formaldehyde gel to a positively charged membrane by capillary transfer. For that purpose, the agarose gel was equilibrated in 20x SSC buffer (3 M sodium chloride, 0.3 M sodium citrate, pH 7) for 2 x 15 min at RT. The membrane was wetted for 1 min with DEPC-H_2O_{bidest} and then soaked together with three Whatman filter papers (similar gel size) in 20x SSC buffer. The blot was assembled as shown in Figure 2.1 and the transfer was performed overnight at 4°C. The blot was quickly washed with DEPC-H_2O_{bidest} to remove salts from the membrane, then the RNA was UV-crosslinked (α = 254 nm) from both sites for 3 min, respectively. Methylene blue staining was performed to visualise the immobilised RNA. The blot was swayed for at most 1 min in 50 ml staining solution containing 50 mg methylene blue, 6.6 ml 3 M

sodium acetate, pH 5.2, 1 ml acetic acid (100 %) and DEPC-H_2O_{bidest}. Destaining was achieved by washing the membrane 3-4 times with DEPC-H_2O_{bidest}. The blot can be stored dry for at least a few months. Electrophoretically separated RNA on a denaturing PAGE (s. 2.5.4.2) was handled the same way, without equilibrating the gel in 20x SSC. The blot was assembled directly after electrophoresis and the membrane and three Whatman filter papers were wetted with 1x TBE.

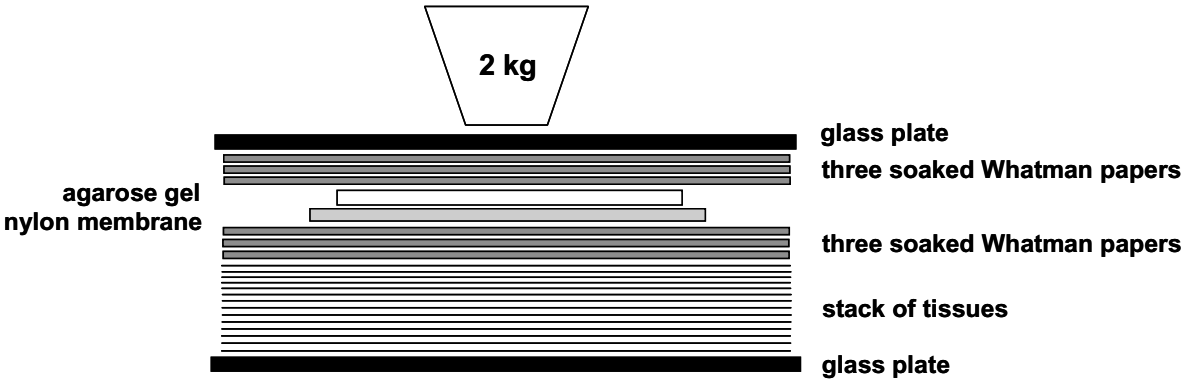


Fig. 2.1: Blot assembly for capillary transfer of RNA from an agarose gel to a positively-charged membrane (Northern blot)

2.5.6 Generation of specific, DIG-labelled RNA probes by *in vitro* transcription

Fragments of selected genes (fragment size ~500 bp), CRISPR substrates (fragment size ~75-2100 bp) and rRNAs (5S and 16S, fragment size ~150-500 bp) were amplified using *Taq* polymerase, purified after gel electrophoresis, quantified photometrically (s. 2.6.7.1, 2.6.6.2, 2.6.4) and finally used as templates for probe generation and labelling by *in vitro* transcription. A sequence-specific forward and a modified reverse primer were used for the amplification. The reverse primer was constructed comprising the T7-binding site and additional six nucleotides for stabilising the T7 polymerase binding (Tab. 2.3).

Reverse primer (5' - 3'):

5' - GGGCCC**TAATACGACTCACTATAGGG** + specific sequence - 3'
6 nt T7

Tab. 2.3: Primer sets of template generation for *in vitro* transcription

Primer	Sequence (5' - 3')	T_m
16S rRNA-for	TTCCGGTTGATCCTGCCGGA	55°C
16S rRNA T7-rev	GGGCCCTAATACGACTCACTATAGGGGGTTACCTTGTTACGACTT	52°C
5S rRNA-for	AAATACCTCTTCAGAGGGACACT	48°C
5S rRNA T7-rev	GGGCCCTAATACGACTCACTATAGGGGCCGCGAGAAAACAGAGCCCT	53°C
TTX4-for	ATTAGCCAGGATCCCTGGAACTTTCA	53°C
TTX4 T7-rev	GGGCCCTAATACGACTCACTATAGGGAACTAAAGACCAGAATTCCCTGAATC	50°C
1rs-for	CAGAATGAGGGATAAGCTTTCA	46°C
1rs T7-rev	GGGCCCTAATACGACTCACTATAGGGAAAGAATTCACGAGGTATCCCA	46°C
cas4-for	CCAGACCTCCCGCGGCATATGTC	57°C
cas4 T7-rev	GGGCCCTAATACGACTCACTATAGGGGGGTTTAGGACCGGCGGCTTC	70°C
csa3-for	GCAGGTGGCTTCATATGAGGATGGGC	58°C
csa3 T7-rev	GGGCCCTAATACGACTCACTATAGGGATTAAGGATCCCTGTCTACAAAGATCCAGCTC	57°C
cas3-for	GGCTGTGTTCCCTGAGGCATATGGTT	56°C
cas3 T7-rev	GGGCCCTAATACGACTCACTATAGGGCATGCTGTACCTTCCCCATC	68°C

In vitro transcription of antisense RNAs was performed by using the T7 *in vitro* transcription kit (Fermentas Life Sciences) according the manufacturer's instructions. Instead of using radioactive labelled nucleotides, the NTP-labelling mixture containing DIG molecules was used (10x: 10 mM ATP, 10 mM CTP, 10 mM GTP, 6.5 mM UTP, 3.5 mM digoxigenin-11-UTP, pH 7.5, Roche). Here, every 20-25th nucleotide of the newly synthesised RNA is a DIG-UTP. A standard reaction assay (20 µl) contained: 500-1,000 ng purified PCR product, 4 µl reaction buffer (5x), 25 U RNase inhibitor, 2 µl NTP-labelling mixture (10x), 20 U T7 RNA polymerase in DNase- and RNase-free H_2O. The samples were mixed, centrifuged briefly and incubated for 2 h at 37°C. The reaction was stopped by the addition of 2 µl 0.2 M EDTA, pH 8. The *in vitro* transcribed RNA was cleaned using LiCl-precipitation (Barlow *et al.*, 1963). The RNA probe was mixed with 0.5 M LiCl and 75 % ethanol incubated at -20°C for 30 min and centrifuged (21,000 x *g*, 20 min, 4°C). Afterwards, the sample was washed twice with 70 % ethanol and in each case centrifuged (21,000 x *g*, 15 min, 4°C). Finally, the RNA was dried and resuspended in 20 µl DEPC-H_2O_{bidest}. The concentration of the RNA probe was determined photometrically (s. 2.5.3). The labelling efficiency was determined by dot blot, as 1 µl of the original reaction sample and 1 µl of a 1:10,

1:100 and 1:1000 dilution were spotted on a nylon membrane, UV-crosslinked and visualised by chemiluminescence (s. 2.5.8). The intensity was compared with a labelled control of known concentration. To ensure, that the RNA probe was not degraded by RNases the homogeneity of the RNA was checked. Therefore, the probe was separated by electrophoresis in an agarose-formaldehyde gel (s. 2.5.4.1), immobilised and detected. The DIG-labelled RNA probe could be stored at -80°C for at least one year.

2.5.7 Hybridisation of immobilised total RNA with DIG-labelled probes

2.5.7.1 Hybridisation of total RNA with antisense RNA probes

1-5 µg of immobilised total RNA (s. 2.5.5) were hybridised with a 500 bp DIG-labelled antisense RNA to detect specific mRNA transcripts in *T. tenax*. First, the blot was prehybridised for 2 h at 68°C in DIG Easy Hyb buffer (Roche) under slight slewing. The prehybridisation step is required to block non-specific binding sites. Before hybridisation, the DIG-labelled RNA probe was denatured by incubation at 95°C for 10 min and afterwards quickly cooled down on ice. The probe was applied to the DIG Easy Hyb buffer in a concentration of 50-100 ng/ml buffer solution and incubated over night at 68°C under slight slewing. After hybridisation, the blot was washed with two stringency solutions, in order to remove unbound probe molecules. First, the blot was washed twice in high-salt buffer (2x SSC, 0.1 % SDS) for 5 min at RT, then washed in low-salt buffer (0.1 % SSC, 0.1 % SDS) 2x 15 min at 68°C. The wet blot was directly used for the immunological detection of RNA hybrids (s. 2.5.8). The hybridisation solutions with the containing probes could be stored at -80°C maximally to one year. For the reuse of the solutions, the RNA probe must be denatured once again for 10 min at 68°C.

2.5.7.2 Hybridisation of small RNA with antisense oligonucleotide probes

1-2 µg of immobilised small RNA (s. 2.5.5) were hybridized with a 43 bp 3'-end DIG-labelled oligonucleotide probe (s. 2.6.11.1). First, the blot was prehybridised for 2 h at 65°C in DIG Easy Hyb buffer (Roche) under slight slewing. The prehybridisation step is required to block non-specific binding sites before hybridisation. Then, the DIG-labelled oligonucleotide probe was directly applied to the DIG Easy Hyb buffer in a concentration of 0.1-0.5 pmol/ml buffer solution and incubated over night at RT under slight slewing. After hybridisation, the blot was washed with a stringency solution, in

order to remove unbound or excessive probe molecules. Therefore, the blot was washed three times with the stringency buffer (2x SSC, 0.1 % SDS) for 5 min at RT, then washed again in the same buffer once for 10 min at 42°C. The wet blot was directly used for the immunological detection of RNA hybrids (s. 2.5.8). The hybridisation solutions were used only once.

2.5.8 Immunological detection of RNA hybrids

RNA hybrids were identified with an alkaline phosphatase-conjugated anti-DIG antibody by a chemiluminescence reaction. Detection of phosphatase activity was carried out by the alkaline phosphatase substrate CDP-Star (Tropix; Roche Manual, 2003; Enger-Blum *et al*., 1993). The blot was equilibrated for 5 min in buffer 1 (0.1 M maleic acid, 3 M sodium chloride pH 8, ad. DEPC-H_2O_{bidest}) at RT, then 2 % (w/v) blocking reagent (Roche) were dissolved in the buffer 1 (buffer 2) and incubation continued for 1 h at RT on a rocking platform to avoid unspecific binding. Anti-digoxigenin-AP (Roche) with a final dilution of 1:20,000 was added to buffer 2 and incubated for 30 min at RT. Subsequently, unbound antibody was washed off using buffer 1 (3x 10 min) and finally the blot was equilibrated in the detection buffer (buffer 3: 0.1 M sodium chloride, 0.1 M Tris/HCl, pH 9.5) for 5 min at RT. For the chemiluminescence reaction the membrane was wetted at the upside with the dye CDP Star (chlorine-substituted 1,2-dioxetan) and placed in a plastic wrap. The chemiluminescence signals were detected using the chemiluminescence detector with an exposure time of 5 min to 4 h.

2.5.9 Nuclease assay

The recombinant protein complexes CasA1 and CasA2 (preparation s. 2.7.2.5, 2.7.2.6) were checked for putative function of degrading different RNAs in the following nuclease assay using DIG labelled RNA fragments as substrates (s. Tab. 2.3): A standard nuclease assay (12 µl volume) consisted of 50 ng DIG-labelled RNA substrates (CRISPR substrates or RNA controls) and 5-500 ng proteins in a chilled reaction buffer (40 mM Tris/HCl pH 7, 10 mM $MgCl_2$, 10 mM ATP, 5 % glycerol, 10 mM ß-Me, 100 mM NaCl). The reaction was started by incubation at 50-65°C for 15 min. For cleaning up the reaction products, an ethanol precipitation with 0.5 M LiCl was performed (s. 2.5.6). The dried RNA was resuspended in 12 µl 10 mM Tris/HCl pH 7 and mixed with 1 vol reaction loading dye (80 % formamide, 2.5 mM EDTA, 0.2 %

bromophenol blue). Before loading on a freshly prepared 1 % agarose-MOPS/formaldehyde (s. 2.5.4.1) or a 12 % urea-polyacrylamide gel (s. 2.5.4.2), the sample was incubated for 10 min at 65°C and quickly chilled on ice. After electrophoresis, the gels were transferred to a positively charged nylon membrane, blotted overnight at 4°C and the RNA UV-crosslinked (s. 2.5.5). The RNA reaction products were identified with an alkaline phosphatase-conjugated anti-DIG antibody by a chemiluminescence reaction (s. 2.5.8).

2.5.10 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays (gel shift assays) represent a technique to evaluate nucleic acid-binding properties of proteins, especially of DNA-binding proteins (s. 2.6.14). With slight modifications also RNA-binding proteins can be studied. The gel shift assay is based upon the fact that the RNA-protein complexes migrate more slowly through a non-denaturing polyacrylamide gel than free RNA.

The DIG labelled RNA fragments (Tab. 2.3), were incubated with a recombinant RNA-binding protein complex of *T. tenax* (CasA2, preparation s. 2.7.2.6). A standard incubation assay (10 µl volume) was composed of 50 ng DIG-labelled RNA and 0.5-150 ng recombinant proteins in binding buffer (40 mM Tris/HCl pH 7, 100 mM NaCl, 10 mM α-Me). The reaction mixture was incubated for 30 min at 37°C, then mixed with 2 µl loading buffer (5x: 0.2 % bromophenol blue, 0.2 % xylene cyanol FF, 60 % (v/v) glycerol, 60 mM EDTA) and directly applied to a freshly prepared 6 % native TBE-polyacrylamide gel (s. 2.6.5.2). After loading the samples onto the gel, the electrophoretic dead time required for the samples to enter the gel matrix was minimised by running the gel at 120 V. Subsequently, the run was continued at 100 V for 90 min. After electrophoresis, the gel was transferred directly to a positively-charged nylon membrane (s. 2.5.5). After blotting, the membrane was briefly washed in DEPC-H_2O_{bidest} and the RNA crosslinked for 2x 3 min at α = 254 nm. Finally, immunological detection of the DIG-labelled RNA was performed (s. 2.5.8).

2.5.11 cDNA synthesis of total RNA from *T. tenax*

With the help of the First Strand cDNA Synthesis Kit (Fermentas) first strand cDNA (complementary DNA) was synthesised from total RNA templates. The used Moloney Murine Leukemia Virus (M-MuLV; Fermentas) reverse transcriptase is a RNA-

directed DNA polymerase, which synthesise a complementary DNA strand initiating from a primer using RNA (cDNA synthesis; McClelland *et al.*, 1993).

The cDNA was produced from previously prepared total RNA (s. 2.5.2.2). First, the prepared RNA was checked for DNA contamination before using in cDNA synthesis. For that purpose, 500 ng of total RNA was applied as a template in a *Taq*-polymerase standard PCR reaction detecting 840 bp of the *T. tenax* hexokinase (TTX_0060, *hxk*) gene (s. 2.6.7.1). As a primer set *hxk* for (5'-TGGTGAGCAGAGATGGGCGAGT-3') and *hxk* rev (5'-ACTTCTTCAGAGTATCCGGCGGC-3') were used with a primer annealing temperature of 60°C (kindly provided by M. Zaparty). Only RNA samples were used in subsequent steps, which gave no correspondent signal in the agarose gel after electrophoresis and ethidium bromide staining. 1-3 µg of total RNA were added to a sterile, nuclease-free tube on ice, mixed with 1 µl random hexamer primer (end concentration: 5 µM) and added DEPC-H_2O_{bidest} to a volume of 11 µl. To dissolve secondary structures the sample was incubated for 5 min at 65°C and quickly chilled on ice. Then, the following components were added: 4 µl reaction buffer (5x conc.), 1 µl RNase Inhibitor (20 U/µl), 2 µl dNTP mix (10 mM), 2 µl M-MuLV Reverse Transcriptase (20 U/µl), gently mixed and incubated for 5 min at 25°C. Because of GC-rich RNA templates, the reaction temperature was increased to 45°C and incubated for 60 min. The reaction was terminated by heating at 70°C for 5 min. The primary cDNA was stored at -80°C or used directly in PCR applications (s. 2.6.7.2).

2.6 Working with DNA

2.6.1 Preparation of genomic DNA from *T. tenax*

For the preparation of genomic *T. tenax* DNA 0.75 g cells were homogenised and lysed in 4 ml DNazol reagent. The preparation was carried out in accordance with the manufacturer's instructions, with slight modifications. The method is based on the use of a guanidinium detergent lysing solution that hydrolyses RNA and promotes the selective precipitation of DNA from the cell lysate (Chomczynski *et al.*, 1997; Mackey *et al.*, 1996). The cells were homogenised by using a hand-held glass-teflon homogeniser. After centrifugation (10,000 x *g*, 10 min, RT) the supernatant was decanted, transferred in a fresh tube and 1 ml ethanol (100%) was added. After

inverting the tube for a few times, the sample was incubated at RT for 3 min. The condensed DNA was spooled, transferred to a fresh tube and mixed with 2 ml ethanol (70 %). The precipitated DNA was pelleted by centrifugation (10,000 x *g*, 10 min, RT) and washed twice with 2 ml ethanol (70 %) and centrifuged again. The remaining ethanol was completely removed under vacuum (speed vac) and the pelleted DNA was then dissolved in 100 µl H_2O_{bidest} and incubated at room temperature for 10 min. To better dissolve the DNA, 1 vol 16 mM NaOH was added. Aliquots of the DNA were stored at -20°C. The DNA preparations were quantified photometrically at α = 260 nm. The quality of purified DNA was checked by restriction enzyme digestion, subsequent agarose gel electrophoresis and the A_{260}/A_{280} ratio (s. 2.6.4).

2.6.2 Preparation of plasmid DNA from *E. coli*

2.6.2.1 Preparation of plasmid DNA by alkaline lysis

Plasmid-DNA isolated by the alkaline lysis method (Birnboim and Doly, 1979, modified) was used for restriction analyses. 2 ml of an overnight culture was centrifuged (12,000 x *g*, 5 min, RT) and the cell pellet was resuspended in 300 µl buffer 1 (50 mM Tris/HCl, 10 mM EDTA pH 8, 100 µg/ml RNase A) followed by the addition of 300 µl freshly prepared buffer 2 (200 mM NaOH, 1% (w/v) SDS) leading to cell lysis during incubation at RT for 5 min. Genomic DNA was precipitated by adding 300 µl chilled buffer 3 (3 M K-acetate pH 4.8), following an incubation on ice for 20 min. After centrifugation (20,000 x *g*, 15 min, 4°C) to remove the genomic DNA, the plasmid DNA was precipitated by the addition of 0.7 vol isopropanol and incubated at RT for 10 min. To recover the precipitated plasmid DNA, the sample was centrifuged (20,000 x *g*, 15 min, 4°C) the supernatant discarded and the plasmid-DNA pellet was washed with ethanol (1 ml, 70 % (v/v) ethanol). The pellet was completely dried under vacuum (speed vac) and finally resuspended in 50 µl H_2O_{bidest}.

2.6.2.2 Plasmid preparations with commercially available kits

Plasmid-DNA used for cloning procedures and sequencing analysis was prepared with the QIAGEN plasmid mini kit (QIAGEN) or the GeneJet™ plasmid miniprep kit (Fermentas) according to the manufacturer's instructions.

2.6.2.3 Preparation of plasmid-DNA for colony PCR

This simple method for the rapid preparation of plasmid-DNA was used to perform a fast screening of recombinant *E. coli* clones. Single colonies of recombinant *E. coli* were picked with a sterilised pipette tip and a portion of the cells were streaked on a LB agar plate. The remaining cells were dissolved in 50 µl 10 mM Tris/HCl, pH 7, incubated at 94°C for 5 min and finally centrifuged (14,000 x *g*, 1 min, RT). 5 µl of the resultant supernatant was used as template for PCR analyses in a 25 µl reaction volume.

2.6.3 DNA precipitation

DNA preparations were concentrated by ethanol or isopropanol precipitation (Sambrook *et al*., 1989). For this, 2 vol 100 % ethanol in presence of 0.1 vol 3 M Na-acetate were added to the sample and incubated for at least 30 min at -20°C or 10 min at -80°C. Alternatively, 0.7-1 vol isopropanol were used. After centrifugation (12,000 x *g*, 10 min, RT) and removal of the supernatant, the DNA pellet was washed with 70 % (v/v) ethanol and again centrifuged. This procedure was repeated, the supernatant finally discarded and the DNA pellet dried under vacuum (speed vac). The pellet was resuspended in an adequate volume of H_2O_{bidest}.

2.6.4 Quantitative and qualitative analysis of DNA

The concentration of the DNA preparations was determined photometrically by measuring the absorption of the sample at α = 260 nm. An absorption of 1.0 (OD_{260} = 1) corresponds to 50 µg of dsDNA/ml or 40 µg of ssDNA/ml (Sambrook *et al*., 1989). The purity of DNA was determined at 260 nm and 280 nm wavelengths. Sufficiently pure DNA preparations showed a ratio of absorbance (A_{260}/A_{280}) of 1.8-2.0. Ratios less than 1.8 indicated that the preparation was contaminated, either with protein or alcohol.

2.6.5 Electrophoresis of DNA

2.6.5.1 Agarose gel electrophoresis of DNA

The electrophoretic separation of DNA molecules by agarose gel electrophoresis (Sambrook *et al*., 1989) was used to determine the size and the amount of DNA molecules (genomic DNA, plasmid DNA, PCR products), to control the results restriction enzyme reactions and to extract DNA fragments for cloning. Depending on the

size of the DNA fragments, agarose gels with 1 % to 1.5 % (w/v) in TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8) were used for electrophoresis. The DNA samples were mixed with loading dye (6x: 0.2 % bromophenol blue, 0.2 % xylene cyanol FF, 60 % (v/v) glycerol, 60 mM EDTA) and applied into the sample wells. 5 µl of a DNA marker containing a mixture of DNA fragments of known size and amount (GeneRuler^{TM}, DNA ladder mix) was also applied onto the gel. Electrophoresis was performed at 80–130 V, depending on the gel size, in TAE-buffer at RT. After the electrophoresis run, the agarose gel was placed for 15 min in a bath with 2 µg/ml ethidium bromide (EtBr). The dye EtBr intercalates in the hydrophobic environment between base pairs of DNA or RNA molecules and emits fluorescence when excited by UV light. The DNA was visualised by exposing the gel to UV light using a Chemi-Doc-gel documentation system. The size and concentration of the separated DNA fragments were determined by comparison of their relative positions to those of the DNA strands of the DNA ladder (Gene Ruler™ DNA Ladder mix, 100-10,000 bp).

2.6.5.2 Non-denaturing polyacrylamide gel electrophoresis of DNA

Electrophoresis separation of DNA, especially for molecules smaller than 200 nt, was also achieved under non-denaturing conditions in polyacrylamide gels (Sambrook *et al.*, 1989). The concentration of the gel (90 mM Tris, 90 mM boric acid, 2 mM EDTA pH 8, 0.03 % (v/v) APS, 0.005 % (v/v) TEMED) varied between 4 and 12 % (v/v) polyacrylamide (acrylamide / bisacrylamide, 40%, ratio 29:1) depending on the size of the desired DNA. The handling and gel cast was performed as mentioned before (s. 2.5.4.2). The DNA samples were mixed with 6x Loading Dye (10 mM Tris/HCl pH 7.6, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 60 % glycerol, 60 mM EDTA) and directly applied on the gel. 5 µl of a DNA marker containing a mixture of DNA fragments of known size and amount (GeneRuler^{TM}, Ultra Low Range DNA ladder, 10-300 bp) was also applied onto the gel. The gel run was performed in 1x TBE (90 mM Tris, 90 mM boric acid, 2 mM EDTA pH 8) at a constant voltage of 80-100 V for 2 h. After the electrophoresis run, the DNA was visualised by EtBr staining.

2.6.6 Purification of DNA fragments

2.6.6.1 Gel extraction of agarose gels

For the extraction and purification of DNA fragments from agarose gels the QIAquick gel extraction kit were used according to the instructions of the manufacturer

(QIAGEN). The DNA was dissolved from the gel matrix by incubation for 10 min at 50°C, bound to a silica gel column under high salt conditions and washed several times. The DNA was eluted under low salt conditions (10 mM Tris/HCl, pH 8.5).

2.6.6.2 Purification of PCR fragments

For the purification of PCR fragments directly from the PCR assay, the QIAquick PCR purification kit or the QIAquick nucleotide removal kit was used according to the instructions of the manufacturer (QIAGEN).

2.6.6.3 Electroelution from polyacrylamide gels

Small DNA fragments were isolated from excised polyacrylamide gel slices by electroelution. The gel slice was placed in a dialysis tube (MEMBRA-Cel®, MWCO 3500, Serva) filled with 2 ml 1x TBE. The dialysis tube was placed across the running direction in an electrophoretic chamber filled with 1x TBE. By an applied voltage of 70 V for 1 h, the DNA migrated from the gel slice into the buffer inside the dialysis tube. Afterwards, the voltage was reversed for 2 min to remove DNA molecules possibly adsorbed to the membrane. The solution was centrifuged (13,000 *x g*, 1 min, RT) and the DNA of the supernatant precipitated with 0.1 vol 3 M Na-acetate and 2 vol 100 % ethanol (s. 2.6.3).

2.6.7 Polymerase chain reactions (PCR)

The PCR technique enables the exponential enzymatic amplification of a specific nucleotide sequence *in vitro* (Mullis *et al*., 1986; Saiki *et al*., 1988). Two sequence-specific oligonucleotide primers hybridise to the 5'-end of the coding and the non-coding strand, respectively, and flank the sequence region, that has to be amplified. These represent the starting points of elongation. The elongation of the primers is catalysed by heat-stable DNA polymerases, e.g. the *Taq*-polymerase of *Thermus aquaticus* (1 min, 1000 bp) or the *Pfu*-polymerase of *Pyrococcus furiosus* (1 min, 500 bp). The latter possesses an additional 3'-5' exonuclease activity ("proofreading-activity"). A standard PCR reaction includes the following main steps:

- Denaturation: The dsDNA is melted on into ssDNA (template) at 94°C
- Primer annealing: Hybridisation of the oligonucleotide-primers to their complementary DNA sequence

- Elongation: DNA polymerase catalyses the elongation of the primers in 5'-3' direction by adding complementary nucleotides

Cycling (25-30x) of denaturation, primer annealing and elongation result in multiple copies of the target sequence. Each primer possesses a specific annealing temperature, which depends on the length and the base composition of the oligonucleotide (n = number of respective bases). The approximate melting temperature (T_m) for primers longer than 13 nucleotides was calculated using the following formula (Wallace *et al*., 1979):

T_m= 64.9 + 41 x (nG+nC-16.4)/(nA+nT+nG+nC)

2.6.7.1 Amplification of genomic DNA and plasmid DNA

The PCR amplifications were performed with 50-100 ng template DNA (genomic or plasmid-DNA), 1 µM of each primer (forward and reverse), 200 µM dNTPs and 1 µl dimethyl sulfoxide (DMSO) in a 25 µl reaction mixture. The reaction buffer (10x) depended on the used DNA polymerase and the required $MgCl_2$ (2 mM, *Taq* polymerase) or $MgSO_4$ (2 mM, *Pfu* polymerase) was already included in the buffers. In general 1 U of the enzyme was used per 25 µl reaction mixture. The PCR reaction was performed using a thermal cycler (BioRad).

Tab. 2.4: Standard PCR assay

	Step 1	Step 2-31	Step 32
Denaturation	94°C, 3 min	94°C, 1 min	
Annealing		40-60°C, 1 min	
Elongation		72°C, 30 sec - 11 min	72°C, 10 min

2.6.7.2 Amplification of primary cDNA

For the amplification of primary cDNA, a modified PCR assay was used. The PCR amplification was performed with 2 µl of synthesised cDNA (s. 2.5.11), 1 µM of sequence specific primer (forward and reverse), 200 µM dNTPs, 5 µl reaction buffer (10x), 2 µl dimethyl sulfoxide (DMSO) and 2 U *Taq* polymerase in 50 µl reaction mixture. The PCR reaction was performed similarly to the standard assay, using a thermal cycler (BioRad). The PCR products were cleaned with the PCR purification kit

(QIAGEN, s. 2.6.6.2) and further analysed by Southern Blots (s. 2.6.12) or by agarose gel electrophoresis (s. 2.6.5.1).

2.6.7.3 PCR mutagenesis

For this PCR amplification, mutagenic primer sets were used to introduce restriction sites at the beginning and the end of DNA fragments. Subsequently, this modified DNA was ligated into adequate vectors (Tab. 2.1). To minimise the error rate, the *Pfu*-DNA polymerase with additional proofreading activity was used.

2.6.7.4 Overlap extension PCR

In the Overlap extension PCR (Ho *et al.*, 1989), two PCR products with overlapping sequences are combined to a longer product, which can be then reamplified. The overlapping sequence contains the mutation site, e.g. to modify an internal restriction site, thus a mutation can be integrated at any position of a PCR product. Two internal (2 and 3) and two external primer (1 and 4) are required. In two separated standard PCR reactions, the primer combinations 1-2 and 3-4 were used to generate the fragments with the designated mutation site in their overlapping sequence. The PCR products were electrophoretically separated and purified (s. 2.6.6.1). In a next step, the two DNA fragments were hybridised (94°C, 3 min) by denaturing both in one sample and afterwards quickly chilled on ice. In an elongation phase for 15 min at 72°C, with the presence of dNTPs and *Pfu*-Polymerase, the complementary DNA strands were completed. Afterwards the combined PCR product was amplified with the external primer set 1 and 4.

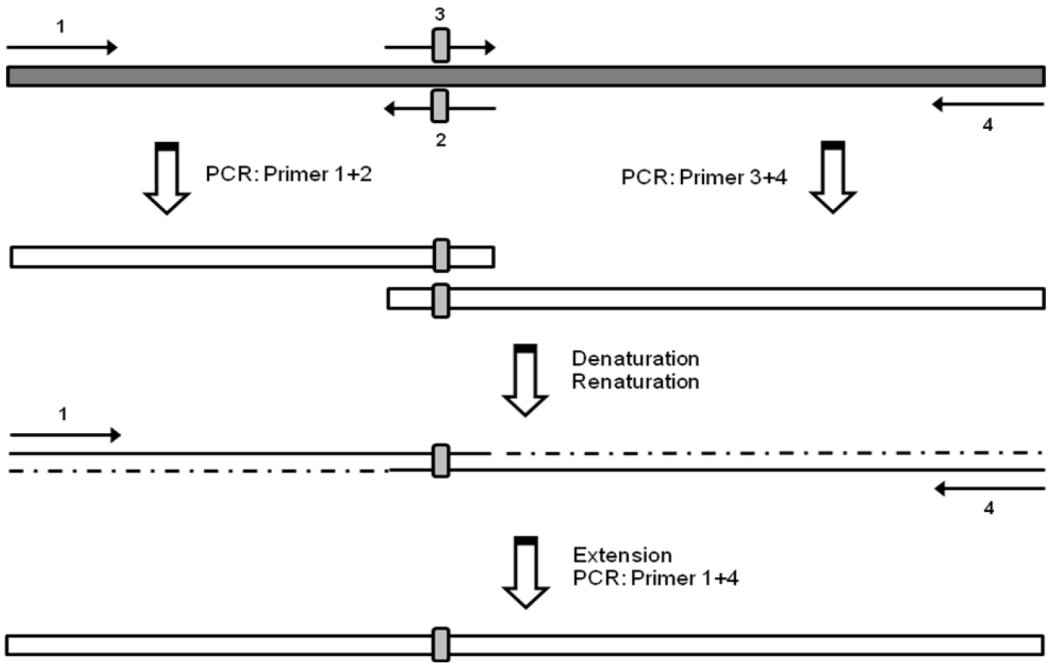


Fig. 2.2: Reaction scheme for the overlap extension PCR

2.6.8 Enzymatic modification of DNA

2.6.8.1 Restriction of DNA

The restriction of genomic DNA, plasmid DNA and PCR products was carried out with respective restriction endonucleases in appropriate buffers following the instructions of the manufacturer. Usually, the reaction batches were incubated at 37°C for 2 h and contained 5-10 U enzyme/µg DNA.

2.6.8.2 5'-dephosphorylation of linearised vector-DNA

In order to avoid self-ligation of restricted vector DNA during the ligation reaction, the 5'-end phosphate groups were eliminated by calf intestine alkaline phosphatase (CIAP) treatment. 0.5 U of CIAP/pmol DNA ends were added directly to the restriction reaction and incubated at 37°C for further 30 min.

2.6.8.3 Ligation

Ligation of restricted DNA fragments (inserts) into vector DNA is carried out by the T4 DNA ligase (Bankier *et al*., 1987; Pan *et al*., 1994). The enzyme catalyses the ATP-dependent formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini of duplex DNA in a blunt-ended or cohesive-ended configuration. Equimolar amounts of restricted, dephosphorylated plasmid DNA and restricted insert were used at a ratio of 1:3 in a volume of up to 16 µl and incubated at 45°C for

5 min to resolve possible secondary structures. 2 µl of 10x reaction buffer (400 mM Tris/HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH 7.8) and 2 µl T4 DNA ligase (1 weiss-unit/µl) were added to the reaction mixture yielding a final volume of 20 µl. The ligation reaction was carried out overnight at 4°C. Subsequently, T4 DNA ligase was inactivated by incubation at 70°C for 10 min. The recombinant vector molecules were stored at -20°C or immediately used for transformation.

2.6.9 Transformation

2.6.9.1 Preparation of competent *E. coli* cells

Competent cells of *E. coli* (s. 2.3) were chemically prepared by using rubidium chloride (RbCl) and calcium chloride ($CaCl_2$, Promega Technical Manual, 1994). 25 ml LB medium was inoculated with 0.25 ml from an overnight culture of *E. coli* and incubated at 37°C in a rotary shaker (200 rpm) until OD_{600} 0.3–0.5 was reached. The culture was then centrifuged (2,300 x *g*, 10 min, 4°C). The cell pellet was gently resuspended in 10 ml chilled solution A (10 mM MOPS, 10 mM RbCl, pH 7.0), centrifuged (2,300 x *g*, 10 min, 4°C), again gently resuspended in 10 ml chilled solution B (100 mM MOPS, 50 mM $CaCl_2$, 10 mM RbCl, pH 6.5) and incubated for 30 min on ice. After the centrifugation (700 x *g*, 15 min, 4°C), the pelleted cells were carefully resuspended in 2 ml chilled solution B and either directly used for transformation (s. 2.6.9.2) or stored with 25 % (v/v) glycerol at -80°C for a few months.

2.6.9.2 Transformation of competent *E. coli* cells

Plasmid DNA was gently mixed with 200-300 µl of competent *E. coli* cells (s. 2.6.9.1) and incubated on ice for at least 1 h. Transformation was achieved by a heat shock at 42°C for 45 sec and subsequent placing on ice for 5 min. 800 µl of LB medium was added and incubated at 37°C for about 1 h in a rotary shaker. 100 µl of transformed cells were plated on LB agar plates containing the respective antibiotics. The remaining 900-1000 µl cell suspension were centrifuged, pelleted cells were resuspended in about 100 µl LB-medium and also plated. After incubation of LB agar plates at 37°C overnight, colonies were screened for positive clones carrying the recombinant plasmid using colony PCR (s. 2.6.2.3) or restriction digestion (s. 2.6.8.1) of isolated plasmid DNA (s. 2.6.2.1, 2.6.2.2).

2.6.10 Sequencing

Automated DNA sequencing (Sanger *et al*., 1977) was performed at AGOWA (Berlin). The following primer sets were used for plasmid sequencing and for internal sequencing of recombinant plasmids:

Tab. 2.5: Vector primer for DNA sequencing reactions

Primer	Sequence (5′ - 3′)	T_m
pET-for	GGATAACAATTCCCCTCTAG	55°C
pET-rev	GCTCAGCGGTGGCAGCAGCC	55°C
casa1-i1	CGTCAAGATAAGGGCTAGG	46°C

2.6.11 3'-end-labelling of oligonucleotides with digoxigenin

Labelling was carried out using the DIG oligonucleotide 3'-end labelling kit (Roche) according to the supplier's instructions. 50-100 pmol of PCR products or oligonucleotides were mixed with 4 µl reaction buffer (5x: 1 M potassium cacodylate, 1.25 mg/ml BSA, 0.125 M Tris/HCl, pH 6.6), 4 µl $CoCl_2$ solution (25 mM), 1 µl DIG-11-ddUTP (1 mM) and 1 µl terminal transferase (400 U/µl) in a total volume of 20 µl. The samples were incubated for 30 min at 37°C and afterwards chilled on ice. The reaction was stopped by the addition of 2 µl of 0.2 M EDTA (pH 8). The DIG-labelled probes were precipitated (s. 2.5.6) or directly stored at -20°C. The labelling efficiency was checked by a spot test, as 1 µl of the original reaction sample and 1 µl of a 1:10 and a 1:100 dilution was spotted on a nylon membrane (s. 2.6.15).

2.6.12 Capillary transfer of DNA to a nylon membrane (Southern blot)

The electrophoretically separated DNA (s. 2.6.5.1) was transferred from the agarose gel to a positively charged membrane by capillary transfer. For that purpose, the agarose gel was first equilibrated for 2x 15 min in a denaturation solution (1 M NaCl, 0.5 M NaOH), then for 2x 15 min in a neutralization solution (3 M NaCl, 0.5 M Tris, pH 7.5) and at last for 10 min in 10x SSC buffer (1.5 M NaCl, 0.15 M Na-Citrate, pH 7). The membrane was wetted for 1 min with H_2O_{bidest} and then soaked together with three Whatman filter papers (similar gel size) in 10x SSC buffer. The blot was assembled as shown in Figure 2.1 and the transfer was performed overnight at 4°C. The blot was then quickly washed with aqua bidest to remove salts from the mem-

brane and the DNA was UV-crosslinked (α = 254 nm) from both sites for 3 min. The blot could be stored dry for few months. Electrophoretic separated DNA on a non-denaturing PAGE (s. 2.6.5.2) was handled the same way, without equilibrating the gel in denaturation and neutralization solutions. The blot was assembled directly after electrophoresis and the membrane and three Whatman filter papers were wetted with 1x TBE.

2.6.13 RT-PCR Southern blot

Immobilised cDNA, amplified by PCR (s. 2.6.7.2) was hybridized with a 500 bp DIG-labelled antisense RNA to verify the specificity of the RT-PCR reaction. First, the blot was prehybridised for 2 h at 52°C in DIG Easy Hyb buffer (Roche) under slight slewing. Before hybridisation, the DIG-labelled RNA probe (s. 2.5.6) was denatured by incubation at 95°C for 10 min and afterwards quickly cooled down on ice. The probe was applied to the DIG Easy Hyb buffer in a concentration of 20 ng/ml buffer solution and incubated over night at 52°C under slight slewing. After hybridisation, the blot was washed with two stringency solutions, in order to remove unbound or excessive probe molecules. First, the blot was washed twice in high-salt buffer (2x SSC, 0.1 % SDS) for 5 min at RT, then washed in low-salt buffer (0.1 % SSC, 0.1 % SDS) 2 x 15 min at 68°C. The wet blot was directly used for the immunological detection of RNA hybrids (s. 2.5.8). The hybridisation solutions with the containing probes could be stored at -80°C not exceeding one year. For the reuse of the solutions, the RNA probe must be denatured once again (10 min at 68°C).

2.6.14 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays (gel shift assays) represent a technique to evaluate DNA-binding properties of a protein and hence to study gene regulation. The principle of an EMSA is based upon the fact that DNA-protein complexes migrate more slowly through a non-denaturing polyacrylamide gel than free DNA.

The PCR amplified and 3'-end DIG labelled DNA fragments (s. 2.6.11.1), which contained putative DNA-binding sites, were incubated with a recombinant putative DNA-binding protein of *T. tenax* (Csa3, preparation s. 2.7.2.3). A standard incubation assay (10 µl volume) was composed of 40 ng DIG-labelled probe and 10-300 ng recombinant protein in binding buffer (50 mM Tris/HCl, pH 8.5, 0.2 µg BSA, 5 % glycerol, 5 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT). Competition experiments were

performed by the addition of 100 ng synthetic poly dI-dC DNA to the reaction mixture. The non-specific Poly dI-dC DNA acted as competitor DNA, ensuring specific DNA-protein interactions. The reaction mixture was incubated for 10 min at 50°C, then mixed with 2 µl loading buffer (6x: 0.2 % bromophenol blue, 0.2 % xylene cyanol FF, 60 % (v/v) glycerol, 60 mM EDTA) and directly applied to a freshly prepared 1 % agarose gel or a 4 % native TBE-polyacrylamide gel (s. 2.6.5.2). After loading the samples onto the gel, the electrophoretic dead time required for the samples to enter the gel matrix was minimised by running the gel at 120 V. Subsequently, the run was continued at 100 V for 90 min. After electrophoresis, the gel was transferred directly to a positively charged nylon membrane, without equilibration in denaturation and neutralisation solutions (s. 2.6.12). After blotting, the membrane was briefly washed in H_2O_{bidest}, the DNA crosslinked and chemiluminescence detected (s. 2.6.15).

2.6.15 Immunological detection of DIG-labelled DNA probes

DIG-labelled DNA probes were identified with an alkaline phosphatase-conjugated anti-DIG antibody by a chemiluminescence reaction with CDP-Star (Tropix) as the alkaline phosphatase substrate. The blot was equilibrated for 5 min in buffer 1 (0.9 % NaCl, 50 mM Tris/HCl, pH 7.4) at RT, then 5 % (w/v) skim milk was dissolved in buffer 1 (buffer 2) and incubation continued for 1 h at RT on a rocking platform to avoid unspecific binding. Anti-DIG-AP (Roche) with a final dilution of 1:20,000 was added to buffer 2 and incubated for 1 h at RT. Subsequently, unbound antibody was washed off using buffer 1 (3 x 15 min) at RT. The chemiluminescence signals were detected with an exposure time lasting from 5 min to 4 h.

2.6.16 *In silico* analysis of nucleotide and amino acid sequences

Software	Reference	Application
CHROMAS 1.3	McCarthy, 1996	Visualisation of sequence chromatograms
FASTA	Pearson & Lipman, 1988	Formatting of nucleotide- and amino acid sequences for submission to databases
BLAST	Altschul *et al*., 1990, 1997 ncbi.nlm.nih.gov/BLAST.cgi	DNA and protein sequence similarity and homology searches
COG database	ncbi.nlm.nih.gov/COG/	Clusters of orthologous groups

Clustal W 2.0	Larkin *et al*., 2007 ebi.ac.uk/Tools/clustalw2/index.html	DNA and protein sequence alignments
IMG 3.0 IMG ER	Markowitz *et al*., 2006, 2009 img.jgi.doe.gov/cgi-bin/pub/main.cgi	Comparative genome analyses
UCSC Archaeal genome browser	Schneider *et al*., 2006 archaea.ucsc.edu/	Comparative genome analyses
BRENDA	Chang *et al*., 2009 brenda-enzymes.org	Enzyme database, comprehensive enzyme information
ExPasy Proteomics Server	Gasteiger *et al*., 2003 expasy.ch	Proteomics server: Analyses of protein sequences, structures and 2-D PAGE
CDD	Marchler-Bauer et al., 2009 NCBI:…/Structure/cdd/cdd.shtml	Conserved domain database
CRISPRfinder, CRISPRdb, CRISPRcompar	Grissa *et al.*, 2007 a/b Grissa *et al.*, 2008 crispr.u-psud.fr	CRISPR database and software for identification and comparison
TIGRFam	Haft *et al.*, 2003 jcvi.org/cms/research/projects/tigrfams/overview/	Database of protein families
Pfam	Finn *et al.*, 2008 pfam.sanger.ac.uk/	Database of protein families
RNAfold	Gruber *et al.*, 2008 rna.tbi.univie.ac.at/	Secondary structure prediction of ssRNA
SMS 2.0	Stothard, 2000 annotathon.univ-mrs.fr/sms2/index.html	Programs for generating, formatting, and analyzing short DNA and protein sequences
Jpred 3	Cole *et al.*, 2008 compbio.dundee.ac.uk/www-jpred/	Secondary protein structure prediction
2Zip	Bornberg-Bauer *et al.*, 1998 2zip.molgen.mpg.de/	Prediction of leucine zippers, leucine repeats and coiled coils
NPS@	Dodd and Egan, 1990 npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth	Helix-turn-helix motif prediction

2.7 Biochemical methods

2.7.1 Heterologous expression of *T. tenax* proteins in *E. coli*

For the heterologous expression of *T. tenax* proteins, the encoding genes were cloned using the T7 polymerase pET vector system (Novagen) via restriction sites introduced by PCR mutagenesis (for nucleotide and amino acid sequences see Appendix, Fig. A4). PCR mutagenesis was performed using *Pfu* polymerase, genomic DNA of *T. tenax* served as the template. The sequences of the cloned genes were verified by dideoxy sequencing (s. 2.6.10). Heterologous expression of the recombinant enzymes was performed in *E. coli* BL21(DE3), BL21(DE3) CodonPlus (pRIL), Rosetta(DE3) and BL21(DE3)pLysS as discussed previously (s. 2.3.2). These *E. coli* strains contain a α prophage (DE3), carrying the T7 RNA polymerase gene and the T7*lac* promoter. In the transformed pET vector constructs, the cloned genes are under control of the T7*lac* promoter and expression is repressed, until IPTG induction of the T7 RNA polymerase from the *lac* promoter. The pLysS plasmid codes for T7 lysozyme, a natural inhibitor of the T7 polymerase, which is constitutively expressed. Therefore, the *E. coli* strains BL21(DE3)pLysS and Rosetta(DE3) were used to minimize low-level expression of gene products before IPTG induction.

2.7.2 Preparation, enrichment and purification of the recombinant enzymes

2.7.2.1 Enrichment of the recombinant Csa5 protein

Recombinant *E. coli* Rosetta(DE3) Csa5 cells were resuspended in chilled buffer 1 (100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$, 300 mM NaCl, 5 ml/g cells) and passed three times through a French pressure cell at 1,100 psi. Cell debris and unbroken cells were removed by ultracentrifugation (45,000 x *g*, 45 min, 4°C). The cell pellet was diluted in 10 mM Tris/HCl pH 7 (1 ml/g) and a 100 µl aliquot kept for subsequent SDS-PAGE analysis at 4°C or at -20°C in Laemmli buffer. For enrichment, the supernatant was subjected to heat precipitation at different temperatures (80-90°C) for 30 min to remove thermolabile *E. coli* proteins. After heat precipitation, the samples were centrifuged (13,000 x *g*, 30 min, 4°C) and dialysed overnight at 4°C against the buffer 2 (100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$). For further purification of Csa5 an anion exchange chromatography with Q-Sepharose® Fast Flow (Amersham) was performed. Therefore, 5 ml resin was packed in a column (diameter: 1 cm), connected to a High Pressure

Liquid Chromatography system (HPLC, Beckman) and equilibrated with buffer 2 (60 min, flow rate: 1 ml/min). The protein solution was loaded onto the equilibrated Q-Sepharose column and was further washed with buffer 2 (20 min, flow rate: 1 ml/min). For releasing the bound proteins, a linear NaCl-gradient was applied (200 min, 0.5 ml/min, buffer 3: 100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$, 1 M NaCl). The run was recorded by following the absorbance at 280 nm. Fractions of 1.5 ml volume were collected consecutively and afterwards analysed on a SDS-PAGE (s. 2.7.4) to follow the purification progress and determine the fractions containing Csa5. The pooled fractions were dialysed overnight at 4°C against buffer 1, concentrated (Vivaspin 20, 10,000 MW) and stored at -20°C in the presence of 25 % glycerol (ultra pure).

2.7.2.2 Enrichment of the recombinant Csa2 protein

Recombinant *E. coli* Rosetta(DE3) Csa2 cells were resuspended in chilled buffer 1 (100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$, 300 mM NaCl, 5 ml/g cells) and passed three times through a French pressure cell at 1,100 psi. The following steps were comparable to the purification of Csa5, except for the heat precipitation step, which was performed only at 80°C for 30 min. For further purification of Csa2 a cation exchange chromatography with Heparin Sepharose™ 6 Fast Flow (Amersham) was performed. Therefore, 5 ml sepharose was packed in a column (diameter: 1 cm), connected to a HPLC system and equilibrated with buffer 2 (60 min, flow rate: 1 ml/min). The protein solution was loaded onto the equilibrated Heparin Sepharose column and the media was further washed with buffer 2 (20 min, flow rate: 1 ml/min). For eluting the bound proteins, a linear NaCl-gradient was applied (200 min, 0.5 ml/min, buffer 3: 100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$, 1 M NaCl). Fractions of 1.5 ml volume were collected and afterwards analysed on a SDS-PAGE (s. 2.7.4). The pooled fractions were dialysed overnight at 4°C against buffer 1, concentrated (Vivaspin 20, 10,000 MW) and stored at -20°C in the presence of 25 % glycerol (ultra pure).

2.7.2.3 Enrichment of the recombinant Csa3 protein

Recombinant *E. coli* Rosetta(DE3) Csa3 cells were resuspended in chilled buffer 1 (100 mM Tris/HCl pH 8.5, 10 mM ß-Me, 5 mM $MgCl_2$, 500 mM NaCl, 5 ml/g cells) and passed three times through a French pressure cell at 1,100 psi. Cell debris and

unbroken cells were removed by ultracentrifugation (45,000 x *g*, 45 min, 4°C). The cell pellet was diluted in 10 mM Tris/HCl pH 7 (1 ml/g) and a 100 µl aliquot was mixed with Laemmli buffer. For enrichment, the supernatant was subjected to heat precipitation at 80°C for 30 min to remove thermolabile *E. coli* proteins. After heat precipitation, the samples were centrifuged (13,000 x *g*, 30 min, 4°C). Since the Csa3 protein was binding a high amount of unspecific *E. coli* DNA, a polyethylenimine (PEI) precipitation was performed. PEI is a strong basic polycation and can be used for the separation of negatively charged nucleic acids and basic proteins (Burgess, 1991; Hardy and Martin, 2008). For this purpose, the heat precipitated protein solution was mixed with 0.5 % PEI (10 % PEI, dialysed in 50 mM Tris/HCl pH 8). The protein solution was incubated on ice and stirred vigorously, while adding dropwise 250 µl PEI. The solution was stirred for additional 30 min and clarified by centrifugation (13,000 *x g*, 30 min, 4°C). The supernatant was precipitated with 80 % ammonium sulphate (AS), for removing residues of PEI, since the proteins are pelleted and the PEI stayed in the supernatant. For the precipitation of the Csa3 protein, the solution was incubated at RT and stirred vigorously, while adding 2.8 g fine pestled AS in portions. Afterwards, the protein was pelleted (13,000 *x g*, 10 min, RT), dissolved in buffer 2 (100 mM Tris/HCl pH 8.5, 10 mM ß-Me, 5 mM $MgCl_2$, 100 mM NaCl) and dialysed overnight at 4°C against buffer 2. For further purification of the Csa3 protein, a cation exchange chromatography with Heparin Sepharose™ 6 Fast Flow (Amersham) was performed. Therefore, 5 ml resin was packed in a column (diameter: 1 cm), connected to a HPLC system and equilibrated with buffer 3 (100 mM Tris/HCl pH 8.5, 10 mM ß-Me, 5 mM $MgCl_2$) for 60 min (flow rate: 1 ml/min). 5 min before sample loading, the NaCl concentration was increased to 100 mM, then the protein solution was loaded and the sepharose was further washed with the same buffer (20 min, flow rate: 1 ml/min). For releasing the bound protein, a linear NaCl-gradient was applied (200 min, 0.5 ml/min, buffer 3: 100 mM Tris/HCl pH 8.5, 10 mM ß-Me, 5 mM $MgCl_2$, 2 M NaCl). Fractions of 1.5 ml volume were collected and analysed on a SDS-PAGE (s. 2.7.4). The pooled fractions were dialysed overnight at 4°C against buffer 2, concentrated (Vivaspin 20, 10,000 MW) and stored at 4°C or used directly in a gel shift assays (s. 2.6.14).

2.7.2.4 Isolation and solubilisation of proteins from inclusion bodies

Due to a very low amount of recombinant protein in the soluble fraction, the proteins were purified from inclusion bodies, which are insoluble aggregates that are often formed during high-level production of recombinant proteins in *E. coli*. Inclusion bodies contain the respective protein in a highly enriched form and can be isolated by *in vitro* refolding techniques (Rudolph and Lilie, 1996; Lilie *et al.*, 1998). For that purpose, 5 g of recombinant Rosetta(DE3) Cas4, Cas1/2, Csa1, Cas5a, Cas3, Cas3HD and Csa4 cells were resuspended in 25 ml buffer 1 (100 mM Tris/HCl pH 7, 1 mM EDTA). Lysozyme (1.5 mg/g cells) was added to the cell suspension, mixed gently and incubated for 30 min at 4°C. For cell disruption, the mixture was passed three times through a French pressure cell at 1,100 psi. After the addition of 10 µg/ml DNase and 3 mM $MgCl_2$, the sample was incubated for 30 min at 25°C. In a next step 0.5 vol of buffer 2 (60 mM EDTA, 6 % Triton X-100, 1.5 M NaCl pH 7) was added and incubated for 30 min at 4°C on a rocking platform. Due to the high density of the inclusion bodies, they can be pelleted by centrifugation (31,000 x *g*, 10 min, 4°C) and therefore, isolated from the prepared lysate. The supernatant was discarded afterwards, the inclusion bodies resuspended in 40 ml buffer 3 (100 mM Tris/HCl pH 7, 20 mM EDTA) and again pelleted by centrifugation (31,000 x *g*, 10 min, 4°C). The obtained inclusion bodies were stored at -20°C maximally two weeks.

For solubilisation of inclusion bodies, 50 mg of the pellet was dissolved in 5 ml buffer 4 (6 M Guanidine hydrochloride (GuHCl), 100 mM Tris/HCl pH 8, 100 mM DTT, 1 mM EDTA) and incubated for 2-4 h at 25°C on a rocking platform. Afterwards, the pH was decreased to 3-4 with 250-300 µl 1M HCl and the sample centrifuged (10,000 x *g*, 10 min, 4°C) to remove cell debris. To eliminate DTT from the sample, two dialysis steps were conducted. First, the protein mixture was dialysed for 2 h at RT against 500 ml buffer 5 (4 M GuHCl, 10 mM HCl), then dialysed overnight at 4°C against 500 ml 4 M GuHCl. The protein concentration was determined by the Bradford protein quantification method (s. 2.7.3). The standard bovine serum albumin was also dissolved in 4 M GuHCl. The solubilised protein could be stored at -80°C for at least one year.

2.7.2.5 Reconstitution of protein complex CasA1

The refolding of the protein complex CasA1 was carried out by rapid dilution. In this case, the denaturing agent GuHCl is removed by a high dilution of the protein solution in a native GuHCl-free buffer (Lilie *et al.*, 1998).

Equal amounts (170 µg) of solubilised protein Cas4, Cas1/2 and Csa1 (s. 2.7.2.4) were added in one sample, mixed vigorously and kept on ice. The proteins were refolded by adding the solution stepwise (every minute, 4-5 µg protein) to 20 ml refolding buffer (40 mM Tris/HCl pH 7, 10 mM ß-Me, 10 % glycerol, 300 mM NaCl, 500 mM L-arginine). Thereby, the refolding assay was incubated at RT and stirred thoroughly. After refolding, the solution was centrifuged (14,000 *x g*, 15 min, 4°C), the supernatant concentrated (Vivaspin 20, 10,000 MW) and used directly for the nuclease assay (s. 2.5.9) or stored at 4°C. The pellet was resuspended in 10 mM Tris/HCl pH 7 and an aliquot mixed with Laemmli buffer for comparing unfolded and refolded CasA1 complexes by SDS-PAGE.

2.7.2.6 Reconstitution of protein complex CasA2

The protein complex CasA2 was refolded by removal of the denaturing agent by stepwise dialysis against a native GuHCl-free buffer (Lilie *et al.*, 1998; Umetsu *et al.*, 2004).

Equal amounts (100 µg) of each solubilised protein Cas5a, Cas3, Cas3HD and Csa4 (s. 2.7.2.4) were combined with 100 µg of the purified proteins Csa5 and Csa2 in one sample. The six proteins were mixed with 5 ml of buffer 1 (3.5 M GuHCl, 2.4 M Urea, 100 mM HEPES/KOH pH 7, 10 % glycerol, 300 mM NaCl, 10 mM $CaCl_2$, 10 mM ß-Me) to receive a protein concentration of 100-120 µg/ml as a starting point for dialysis. Furthermore, 30 µg total *T. tenax* RNA preparated by isopropanol precipitation (s. 2.5.2.1) was added to the protein-buffer mixture for supporting the refolding process of CasA2. The solution was mixed thoroughly, filled in a dialysis bag (volume: 6 ml) and dialysed at RT against 500 ml of buffer 2 (3 M GuHCl, 2 M Urea, 100 mM HEPES/KOH pH 7, 10 % glycerol, 300 mM NaCl, 10 mM $CaCl_2$, 10 mM ß-Me).

After dialysis, the protein sample was centrifuged (14,000 *x g*, 15 min, 4°C), the supernatant concentrated (Vivaspin 20, 10,000 MW) and used directly in gel filtration chromatography (s. 2.7.2.5) or stored at 4°C. The pellet was resuspended in 10 mM Tris/HCl pH 7 and an aliquot mixed with Laemmli buffer for comparison of unfolded and refolded CasA2 complexes on a SDS-PAGE.

Tab. 2.6: Stepwise dialysis protocol for refolding of complex CasA2

Step	M GuHCl	M Urea	Time, h
A	3	2	2
B	2	2	2
C	1	2	2
D	0	2	1
E	0	1	1
F	0	0	0.5

2.7.2.7 Purification of His-tagged recombinant enzymes

The *T. tenax* genes Cas4, Csa1 and Cas3 were cloned in frame with a C-terminal His-tag (six codons: CAC), adding six histidine residues to the C-terminus of the protein, in order to purify and enrich recombinant proteins after heat precipitation from the soluble fraction. His-tag specific affinity chromatography was performed using Ni-NTA agarose.

The *E. coli* Rosetta(DE3) cells were resuspended in chilled buffer 1 (20 mM Tris/HCl pH 8, 10 % glycerol, 10 mM ß-Me, 1 mM EDTA, 100 mM NaCl, 5 ml/g cells) and handled the same way as above-mentioned (s. 2.7.2.1). The supernatant was heat precipitated at 70°C for 30 min, centrifuged (13,000 x *g*, 30 min, 4°C) and dialysed overnight at 4°C against the buffer 1. 4 ml Ni-NTA agarose was equilibrated with 50 ml buffer 2 (20 mM Tris/HCl pH 8, 10 % glycerol, 10 mM ß-Me, 300 mM NaCl). After washing the agarose (30 ml, buffer 1), elution was performed with 20 ml buffer 2 (buffer 1: containing 250 mM imidazole). Fractions of 1.5 ml were collected and subsequently analysed on a SDS-PAGE (s. 2.7.4). The pooled fractions were dialysed overnight at 4°C against buffer 2, concentrated (Vivaspin 20, 10,000 MW) and stored at 4°C.

2.7.3 Protein quantitation

The determination of protein concentration was carried out using Bio-Rad Protein Assay based on the Bradford protein quantitation method (Bradford, 1976; modified) and following the instructions of the supplier. Bovine serum albumin (BSA; 2-10 µg/ml) served as standard.

2.7.4 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

2.7.4.1 SDS-PAGE assembling and electrophoresis

For protein analyses, denaturing sodium dodecylsulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was applied (Laemmli, 1970). Polyacrylamide gels (8.6 cm x 7.7 cm, 1 mm thick) were used as support matrix in electrophoresis, composed of an upper stacking and a lower separating gel. The acrylamide concentration of the stacking gel (125 mM Tris/HCl, pH 6.8, 0.1 % (v/v) SDS, 0.03 % (v/v) APS, 0.005 % (v/v) TEMED) was 4.9 % (v/v) and the concentration of the separation gel (375 mM Tris/HCl, pH 8.8, 0.1 % (v/v) SDS, 0.03 % (v/v) APS, 0.005 % (v/v) TEMED) varied between 10 and 15 % (v/v) polyacrylamide due to the molecular weight range of the protein subunits as well as the desired separation of the proteins. Ingredients for the separating gel were mixed and poured in the gel casting chamber. The gel was covered with butan-1-ol and polymerised for 20 min. The butan-1-ol was then removed, the gel surface was washed with H_2O_{bidest}, dried and the stacking gel solution was poured upon. A 10-sample well comb was placed into the stacking gel and removed after approximately 20 min. The gels were either directly used for electrophoresis or stored at 4°C. Prior to electrophoresis, the protein samples were mixed with 5x loading buffer (final concentration: 62.5 mM Tris-HCl, pH 6.8, 10 % (v/v) glycerol, 2 % (v/v) SDS, 5 % (v/v) α-Me, 0.005 % (w/v) bromophenol blue) and heated for 2-3 minutes at 94°C. Gel runs were performed in a Minigel-Twin-Chamber (Biometra) containing electrophoresis buffer (25 mM Tris-HCl pH 8, 190 mM glycine and 0.1 % (v/v) SDS) at 13 mA for 30 min until the dye front reached the separating gel and continued at 17 mA for 90 min.

2.7.4.2 Coomassie staining

Proteins were visualised by gel staining (40 % (v/v) methanol, 10 % (v/v) acetic acid and 0.25 % (w/v) Coomassie Brilliant Blue R-250) for 30 min at 50°C and destaining (5 % (v/v) methanol and 7.5 % (v/v) acetic acid) (Weber and Osborn, 1969). The gels were documented using the ChemiDoc System in combination with Quantity One Software Package (BioRad).

2.7.4.3 Molecular mass determination of proteins under denaturing conditions

The approximate molecular mass of protein subunits was determined by the migration distance in a SDS-PAGE comparing to a standard protein mixture (PageRuler™

Unstained Protein Ladder). This marker is a mixture of 14 recombinant, highly purified, unstained proteins ranging in size from 10 kDa to 200 kDa. The 50 kDa protein has a clearly greater intensity and served as a reference band.

2.7.5 Molecular mass determination of proteins under native conditions

The approximate molecular mass of protein complexes was determined by gel filtration with Superose™ 6 10/300 columns (volume: 24 ml, length: 30 cm, diameter: 0.5 cm, Pharmacia). The gel filtration was performed with a HPLC in a buffered solution (100 mM HEPES/KOH pH 7, 10 % glycerol, 10 mM ß-Me, 10 mM $CaCl_2$, 300 mM NaCl) and the resin was equilibrated (12 h, flow rate: 0.3 ml/min) prior to sample loading. 200 µl of the prepared protein solution (s. 2.7.2.6) was loaded and 0.45 ml fractions were collected (flow rate: 0.3 ml/min). 400 µl of the fractions were precipitated with 4 vol of chilled 100 % acetone, incubated for 30 min at -20°C, centrifuged (13,000 *x g*, 30 min, 4°C), dried and resuspended in 20 µl 10 mM Tris/HCl pH 7. The samples were further analysed by SDS-PAGE (s. 2.7.4).

As reference proteins Ferritin (MW 443,000, 1 mg), Alcohol-Dehydrogenase (MW 148,000, 1.25 mg), D-Lactate-Dehydrogenase (MW 78,000, 0.118 mg) and Cytochrome C (MW 12,500, 2 mg) were used.

- Ferritin: specific absorption at 217 nm
- Alcohol-Dehydrogenase: activity measurement at 340 nm (RT) in 0.1 M Tris/HCl pH 7, 0.4 mM NADH, 2 mM acetaldehyde
- D-Lactate-Dehydrogenase: activity measurement at 340 nm (RT) in 0.1 M Tris/HCl pH 7, 0.4 mM NADH, 2 mM pyruvate
- Cytochrome C: specific absorption at 416 nm

2.7.6 Determination of the amino acid sequence

The determination of the amino acid sequence of electrophoretically separated proteins was carried out by trypsin proteolysis and MALDI-TOF analyses (kindly performed by E. Jeske, Institut für Hygiene und Arbeitsmedizin, Universitätsklinikum Essen).

3. RESULTS

3.1 CRISPR loci in *T. tenax*

The impact of clustered regularly interspaced short palindromic repeats (CRISPR) on the prokaryotic cell is only partially understood until today. Recently, it became clear that CRISPR are a common feature of prokaryotes, since 94 % of all archaeal and 44 % of all bacterial strains possess such clusters within their genomes (resource CRISPRdb, date: 12-2009; Grissa *et al.*, 2007). They are characterised by direct repeats with a size ranging from 24 to 48 base pairs and separated by slightly longer spacer sequences. CRISPR loci are generally flanked up- or downstream by an AT-rich leader sequence of 200-350 bp length, supposed to act as a promoter for RNA transcription of the CRISPR locus (Jansen *et al.*, 2002).

3.1.1 Identification and features of *T. tenax* CRISPR

Until this work, de facto nothing was known about the occurrence of CRISPR in *T. tenax*. For identifying CRISPR loci of *T. tenax,* the genome of this organism was screened with the help of the CRISPRFinder web tool (Grissa *et al.*, 2007). This programme uses following criteria for their identification: (i) Multiple repeats exhibit a length of 23 to 55 bp and a gap size of 25 to 60 bp with one nucleotide mismatch between the repeats is accepted; (ii) length of the spacers ranges between 0.6 and 2.5-fold of the repeat length; (iii) the individual spacer sequences are not identical. By applying these criteria seven CRISPR loci were proposed termed TTX_1 to TTX_7, characterized by a repeat length of 24-25 bp and a number of 6 to 33 spacers (Tab. 3.1, sequences Appendix A1). Based on preferred sequence similarities of the repeat elements, the seven CRISPR loci were separated into two groups. Group I cluster (TTX_1, TTX_4, TTX_5, TTX_6 and TTX_7) show a higher number of repeat/spacer units (15-33 vs. 6-7 units) and a greater diversity of spacer length (37-57 vs. 42-49 bp) as the group II cluster (TTX_2 and TTX_3). All repeat sequences show some dyad symmetry, but are not truly palindromic and this observation is in line with previous analysed CRISPR loci from numerous prokaryotes (Jansen *et al.*, 2002). In order to identify the leader sequence and with it the correct orientation of the CRISPR in the genome, both flanking sites (269-406 bp) of each cluster were compared with respect to their AT-content (SMS, DNA Stats; Stothard, 2000). The flanking sites with the higher AT-content (4.78 % (± 2.8 %)) compared to the opposite flanking sites has

been labelled as the leader sequence defining the orientation of each CRISPR locus. Remarkably, in two cases (TTX_4 and TTX_6) a degenerated repeat was localised at the opposite side of the leader sequence (Siebers *et al.*, 2010; submitted).

Tab. 3.1: The seven CRISPR loci of *T. tenax*. The complete genome of *T. tenax* was analysed via CRISPRFinder; the orientation (Ori) is defined by the localisation of the leader sequence.

CRISPR	Position	Repeat elements	Ori	Group	Spacer	Length
TTX_1	224864-7260	GAATCTCAGATAGAGATTTGAAGG	rev	I	24	39-50 nt
TTX_2	316783-7791	GTGGAAATCAAAAGATAGTAGAAAC	for	II	6	42-46 nt
TTX_3	345073-6121	GTGGAAATCAAAAGATAGTAGAAA	rev	II	7	42-49 nt
TTX_4	1075638-8562	GAATCTCAAAGAGAGGATTGAAAG	rev	I	33	37-57 nt
TTX_5	1081524-4202	GAATCTCAAAGAGAGGATTGAAAG	for	I	31	38-51 nt
TTX_6	1094882-7343	GAATCTCAAAAAGAGGATTGAAAG	rev	I	26	37-51 nt
TTX_7	1102723-4458	GAATCTCAAAGAGAGGATTGAAAG	for	I	15	39-55 nt

Taken together, 143 unique spacer sequences have been detected within the seven CRISPR loci. The distribution of the spacer length showed that the highest number of spacers had a length of 43 bp and that nearly 75% of all spacers were in the range between 41 and 46 bp (Fig. 3.1). So, it seems that there exists a preference for a distinct length of the spacer sequences.

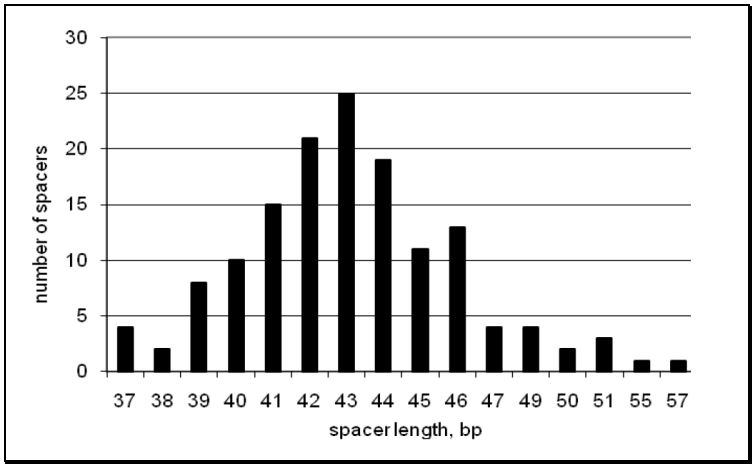


Fig. 3.1: Distribution of the length of spacer sequences in the seven CRISPR loci of *T. tenax*.

3.1.2 Assignment of all spacers of the seven CRISPR loci

An actual hypothesis about the origin and function of CRISPR bases on the observation that some spacer sequences showed similarity to parts of extra-chromosomal elements, such as plasmids, transposons or viruses (Mojica *et al*., 2005; Pourcel *et al*., 2005; Bolotin *et al*., 2005). So it was tried to distinguish the origin and similarity of the identified spacer sequences in *T. tenax*. Due to the fact that leader sequences mark the beginning of a CRISPR array, all spacers got an identification number (Fig. 3.2, Appendix A2).

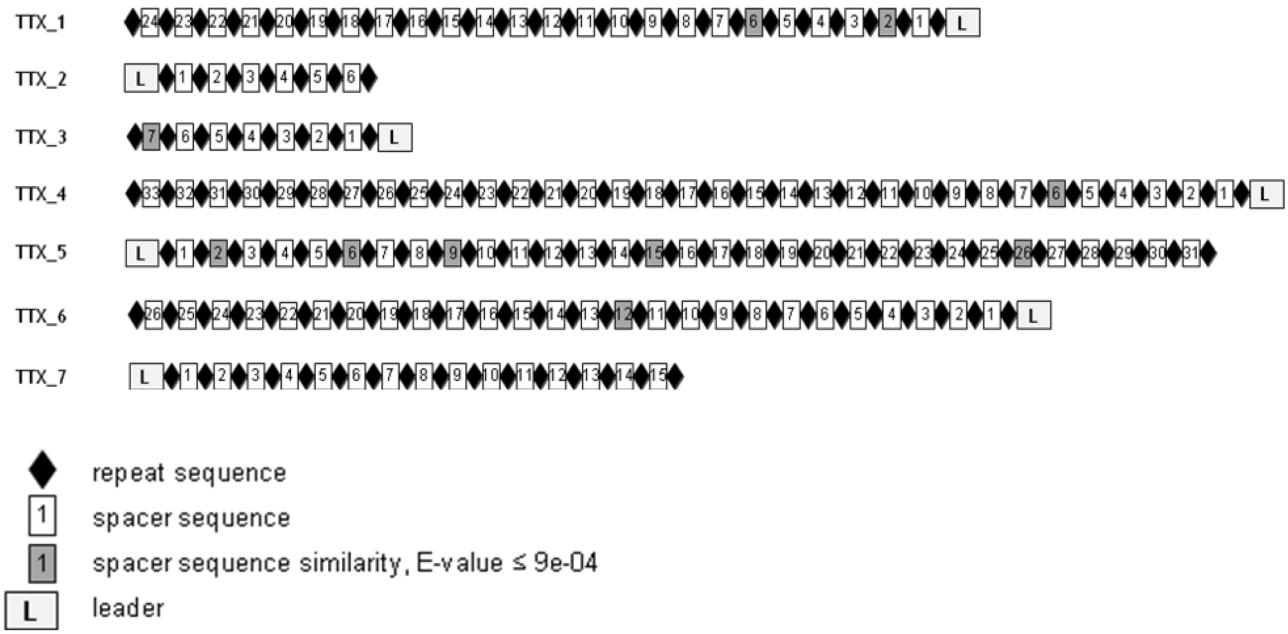


Fig. 3.2: Numbering of spacers of the seven CRISPR loci. All identified spacers of the seven CRISPR loci are consecutively numbered starting from the leader sequence site (example: spacer 1.1, spacer 1 from TTX_1).

Various databases were applied for the homology searches of spacer sequences (s. 2.6.16). First, the BLASTn tool (National Centre of Biotechnology NCBI; Altschul *et al.*, 1990) was used for a first, general scanning for similarities between spacers and the nucleotide database (Appendix A2, word size: 7, E-value: 0.02). However, no significant hits could be detected. In a next step, all spacers from *T. tenax* were compared to the spacer sequences from 467 prokaryotic organisms in the CRISPRdb (E-value: 0.1). The highest similarities were only found for a few spacer sequences of closer related archaeal organisms (*Thermofilum pendens*, *Hyperthermus butylicus*). In another approach, the *T. tenax* spacer sequences were not only compared to CRISPR spacers of various prokaryotes, but also to the whole genome sequences of

T. tenax and of close related organisms (*T. pendens*, *Pyrobaculum aerophilum*, *Thermoproteus neutrophilus*; Fitz-Gibbon *et al.*, 2002; Anderson *et al.*, 2008). In Table 3.2 ten hits with the lowest E-values are listed. Interestingly, the highest similarity was detected between spacer 4.6 and the reading frame TTX_0660 of *T. tenax*, which putatively encodes an adhesin-like protein with transmembrane-helices motifs. A preferred similarity was also found between spacer 5.6 and the putative coding gene for the cell division control protein Cdc21 (TTX_0274) and between spacer 5.9 and the putative coding gene of a creatinine amidohydrolase (TTX_0331). At last, all *T. tenax* spacers were checked against the genome sequences of 31 viruses infecting archaeal organisms, i.a. *Thermoproteus tenax virus 1* (TTV1), *Thermoproteus tenax spherical virus 1* (TTSV1) and *Pyrobaculum spherical virus* (PSV), which are known to infect *T. tenax*. Hits with the highest similarity were detected for genes of *Acidianus two-tailed virus* (ATV), *Acidianus filamentous virus 6* (AFV6), PSV or *Sulfolobus islandicus rod-shaped virus 2* (SIRV2). However, the origin of a vast majority of spacer sequences in the *T. tenax* genome are unknown.

Tab. 3.2: Preferred sequence similarities of *T. tenax* CRISPR spacers. All 143 spacer sequences were analysed with BLASTn, CRISPRdb and IMG. Listed are ten hits with the lowest E-value (complete results and spacer sequences, Appendix A2).

Spacer	Genome	Hit	E-value
4.6	*Thermoproteus tenax*	TTX 0660_adhesin-like transmembr.-helices	3e-15
1.2	*Acidianus two-tailed virus*	gp09_conserved hypothetical	1e-04
1.6	*Thermofilum pendens*	Tpen 1516_alcohol dehydrogenase	2e-04
5.6	*Thermoproteus tenax*	TTX 0274_cell division control protein Cdc21	2e-04
5.9	*Thermoproteus tenax*	TTX 0331_creatinine amidohydrolase	2e-04
6.12	*Acidianus filamentous virus 6*	gp31_conserved hypothetical	2e-04
5.2	*Pyrobaculum spherical virus*	gp28_conserved hypothetical	7e-04
5.15	*Acidianus rod-shaped virus 1*	gp23_glycosyl transferase	7e-04
3.7	*Thermofilum pendens*	Tpen 0049_conserved hypothetical	7e-04
1.2	*Sulfolobus islandicus r.-s. virus 2*	gp19_CRISPR-associated protein Cas4	9e-04

3.1.3 Analyses of small RNAs in *T. tenax* by Northern blots

As deduced from studies of other organisms, CRISPR loci are transcribed into long RNA transcripts, potentially covering the entire cluster and then step-wise processed into small crRNAs, generating transcripts with a length of 60-70 nt (Tang *et al.*, 2002 and 2005; Hale *et al.*, 2008). The analyses should give indications for an stress induced transcription of the CRISPR loci in *T. tenax*.

To study the transcription of the seven CRISPR loci of *T. tenax* in response to (abiotic) stress, Northern blot analyses of small RNA fractions from stressed *T. tenax* cultures were performed using "normally" grown *T. tenax* cultures (optimal heterotrophic conditions at 86°C and pH 5.5) as a control. For stress induction, the cultures were irradiated by UV-light (for 30 sec or 2 min) or incubated at increased temperature (91°C, for 3 h; see 2.3.1.2 and 2.3.1.3). RNA was prepared from 0.1 g *T. tenax* cells of the late exponential growth phase by the use of the *mir*Vana™ miRNA Isolation Kit (Ambion), ending up in two 100 µl RNA fractions, containing small (α 200 nt) and large RNAs (α 200 nt) species (s. 2.5.2.3). In quantitative photometric measurements, both RNA fractions provided reliable amounts for all tested cells (s. 2.5.3 and Tab. 3.3).

Tab. 3.3: Quantitative analysis of RNA preparations isolated with *mir*Vana™ miRNA Isolation Kit. In each case (control, 91°C upshift, UV-light stress), RNA was prepared from 0.1 g *T. tenax* cells. Averaged concentrations and standard deviations of three independent preparations (derived from three independent cultivations) are listed; **a)** small RNA fraction (10-200 nt); **b)** large RNA fraction (α 200 nt)

a) small RNA

Cells	ng/µl	A_{260}/A_{280}	A_{260}
control	116 (± 54)	1.88 (± 0.10)	0.097 (± 0.034)
91°C	106 (± 44)	1.82 (± 0.05)	0.080 (± 0.033)
UV30"	233 (± 48)	1.80 (± 0.05)	0.176 (± 0.037)
UV120"	168 (± 66)	1.82 (± 0.08)	0.127 (± 0.050)

b) large RNA

Cells	ng/µl	A_{260}/A_{280}	A_{260}
control	330 (± 140)	1.96 (± 0.07)	0.244 (± 0.087)
91°C	369 (± 141)	1.86 (± 0.10)	0.236 (± 0.078)
UV30"	609 (± 126)	1.86 (± 0.06)	0.381 (± 0.079)
UV120"	534 (± 311)	1.91 (± 0.07)	0.334 (± 0.194)

In addition to quantitative, also qualitative analyses of prepared RNA species were performed by using agarose gel electrophoresis. For this purpose, 1 µg small RNA of each cell type was electrophoretically separated on a 12 % denaturing PAGE (s. 2.5.4.1) and 2 µg of the large RNA fractions on a 1 % denaturing agarose gel (s. 2.5.4.2), respectively (Fig. 3.3). The Northern blots revealed a satisfying integrity of both prepared RNA fractions. The signals of the 5S rRNA and tRNAs (small RNA), and the 23S and 16S rRNA (large RNA) served as an indicator for the integrity of the applied RNA.

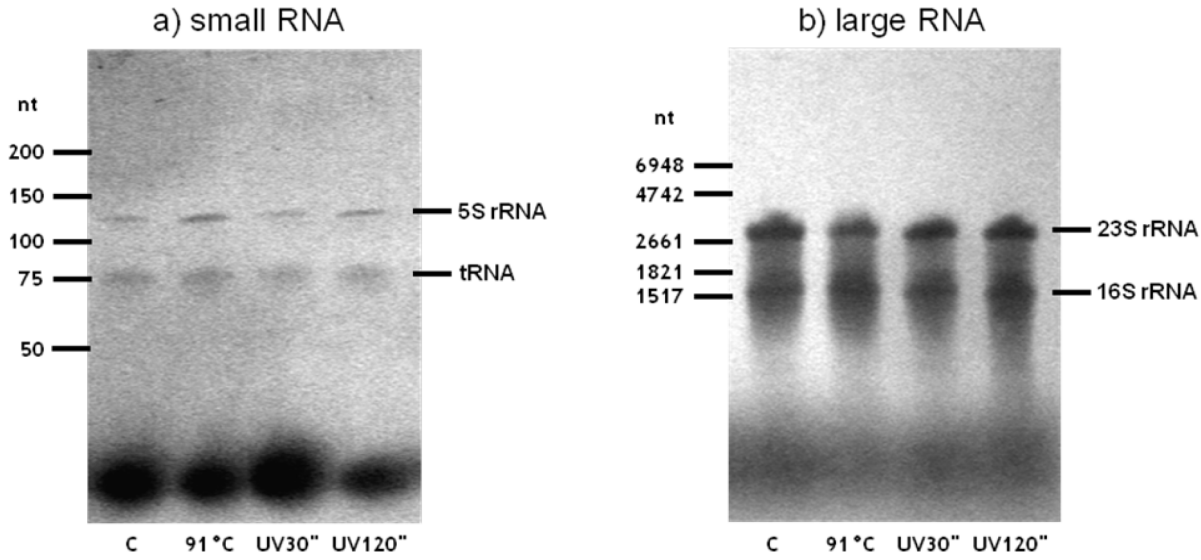


Fig. 3.3: Methylene blue stained Northern blots of separated small and large RNA. a) 1 µg small RNA of *T. tenax* control cells (C) and stressed cells (91°C upshift, 30" and 120" UV-light irradiation), 12 % denaturing PAGE, marker: Ultra Low Range DNA ladder. **b)** 2 µg large RNA of *T. tenax* control cells (C) and stressed cells, 1 % denaturing agarose gel, marker: RNA Molecular Weight Ladder II.

For the Northern blot analyses of CRISPR RNA transcripts 1 µg of prepared small RNA species of different *T. tenax* cell types were electrophoretically separated on a 12 % denaturing PAGE and blotted (s. 2.5.4.2 and 2.5.5). For identification with CRISPR-specific nucleotide probes, from each CRISPR locus a spacer sequence was selected, which showed no similarity to the rest of the *T. tenax* genome to avoid unspecific probe hybridisation. Moreover, all oligonucleotide antisense probes had a length of 43 bp to facilitate the comparison of results (s. 2.6.11.1 and Tab. 3.4). For the hybridisation of immobilised small RNA with the DIG-labelled oligonucleotides, a protocol based on the Ambion instruction's manual was used (s. 2.5.7.2). The detection of hybrids was carried out by chemiluminescence (s. 2.5.8). The 5S rRNA served as an internal standard to ensure equal amounts of small RNA have been applied.

Tab. 3.4: Antisense spacer probes for Northern blot analyses of the seven CRISPR loci. From each CRISPR locus in *T. tenax* one spacer sequence with a length of 43 bp was selected; antisense oligonucleotides to selected spacer sequences were labelled with DIG Oligonucleotide 3'-End Labelling Kit and used as probes in Northern Blot analyses.

Probe	Sequence (5' - 3')	Length, bp	Ori CRISPR
1.20 for	AGATCCATGTGCCGATCTTCGTGTTGCCCGTCTCGTCGAATAT	43	rev
2.6 rev	ACCACGCGAGAAAAATACATAGCATGGGGTGCCGATGTGTATG	43	for
3.5 for	TCAATGGCGATACAGCGTGGGGGCCTGTAGTGCCATGGCTCAC	43	rev
4.22 for	AGATTCGAGCTCTTTGCTGGCAGATTGGGATACCTCGAGAAGA	43	rev
5.14 rev	CATATAGCCGACACGCCCGCATTCTGCCCATCGATGATCCAGT	43	for
6.15 for	AAGAGGTAGTTAAGCACTTCAACGCAATAGGCTTTATGCCCTG	43	rev
7.8 rev	TCGATGAGATCAGATGTCCATATTGCTTCCGCCGTGTCTATCC	43	for

As shown in Fig.3.4, the Northern blot analyses identified small RNA transcripts from five of the seven CRISPR loci. Interestingly, the detected small RNAs showed different sizes (~130 nt, ~110 nt, ~70 nt, ~50 nt), suggesting that large CRISPR transcripts were processed step-wise to small RNAs. But even the smallest detectable transcripts of ~50 nt were larger then the simple spacer sequences of 43 bp, which implies that the RNA transcripts should be flanked by the adjacent repeat(s), either by the whole or partial repeat sequence(s). For the CRISPR loci

TTX_2 and 3, which are designated as type II CRISPR, no signals appeared on the blot, even after an exposure time up to 4 h. However, no differences were detected between the control and stressed cells of *T. tenax*. Obviously, the generation of small CRISPR RNAs is not induced under the abiotic stress conditions tested.

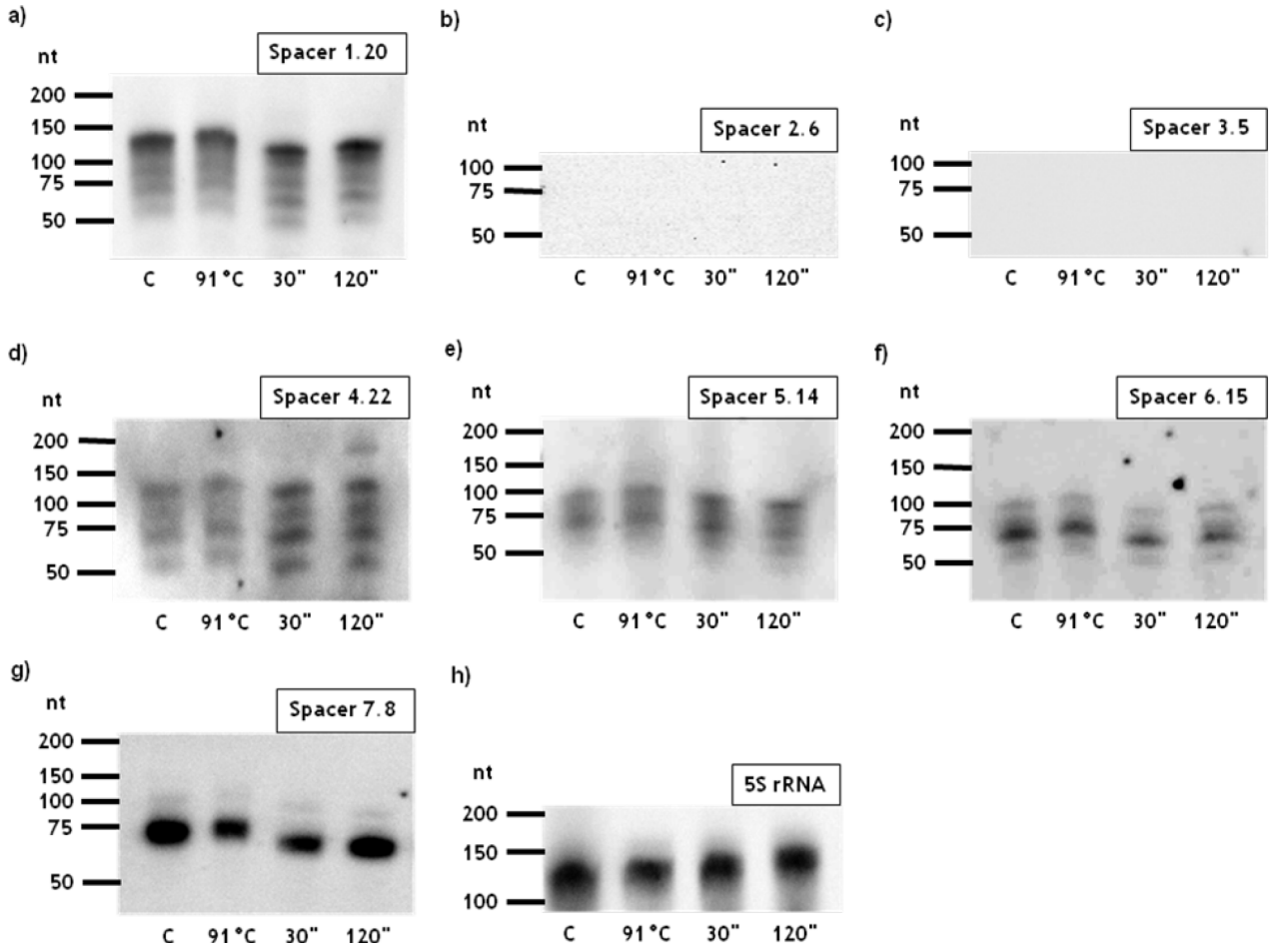


Fig. 3.4: Northern blot analyses of small RNA from the seven CRISPR loci of *T. tenax*. Hybridisation of DIG-labelled antisense spacer sequences with 1 µg immobilised small RNA (α 200 nt). RNA from control (C), 91°C upshift (91°C), 30 (30") and 120 (120") sec UV-light stressed *T. tenax* cells used. **a)** + **d-g)** CRISPR locus TTX_1, TTX_4, TTX_5, TTX_6 and TTX_7 of type I. **b-c)** CRISPR loci TTX_2 and TTX_3 of type II. **h)** 5S rRNA internal standard

Remarkably, for the type II CRISPR loci (TTX_2, TTX_3) characterised by their lower number of repeat/spacer sequences and the lower variability of the spacer sequences, no small RNA species were detectable, not even under control conditions. Possibly, this observation hints for a lower stability of the respective RNA species. To substantiate this suggestion, we focussed at the formation of possibly stabilising structure elements in the repeat sequences involved in the formation of the small

RNA, since Kunin and co-workers (2007) reported that especially these elements determine the stability of the deduced RNA species. Therefore, the repeat sequences of the seven CRISPR loci were analysed with the RNAfold tool to compare their predicted secondary RNA structure. As shown in Fig. 3.5, the prediction for type I loci is a stable stem-loop structure (TTX_1: 6.38 kcal/mol; TTX_4 to TTX_7: 3.22 kcal/mol). For type II loci, however, the formation of such a stabilising secondary structure seems to be improbable (TTX_2 and TTX_3: 0.52 kcal/mol). Possibly, the low stability of TTX_2 and TTX_3 is the cause for the absence of respective signals of small RNA in the Northern blot analyses.

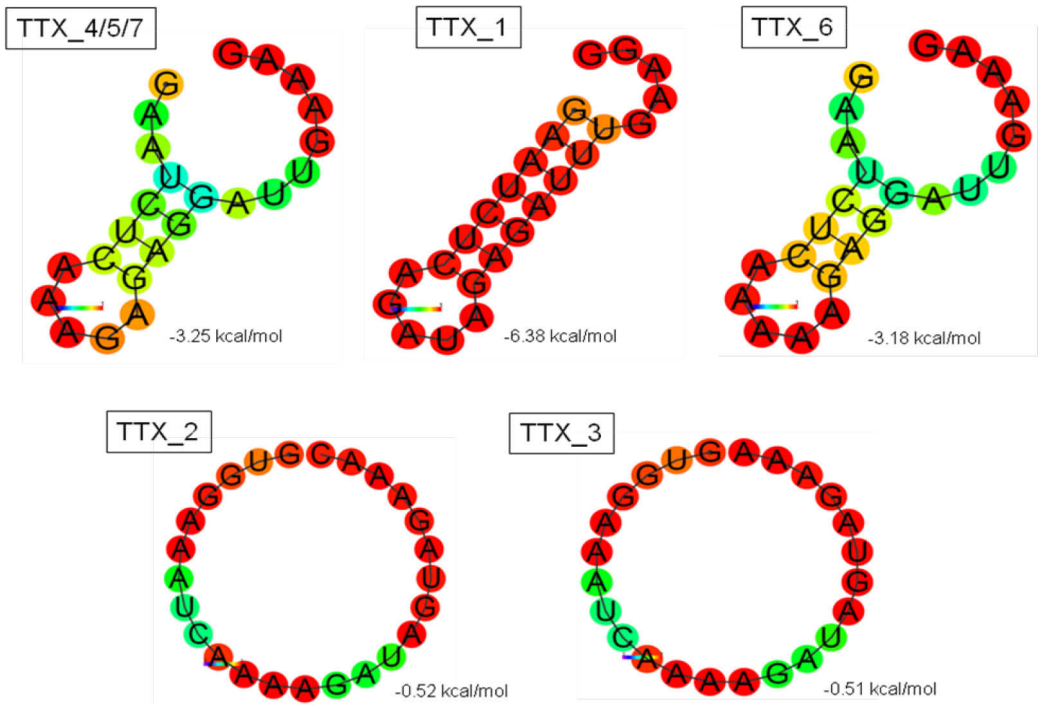


Fig. 3.5: Secondary structure prediction of repeat sequences. Repeats of the CRISPR were analysed with the RNAfold tool. TTX_4/5/7 had the same repeat sequence. TTX_1, 4-7 belong to type I, TTX_2-3 to type II cluster. Depicted are the minimum free energy structures and the free energy values of the ensembles.

3.2 Identification and transcription of *T. tenax cas* genes

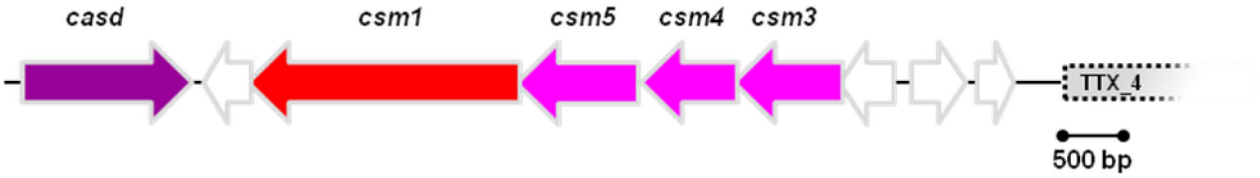
Adjacent to the CRISPR loci, a set of conserved *CRISPR-associated* (*cas*) genes are located, completing the CRISPR/Cas system in *T. tenax*. But as comparative genomics revealed, there is a huge quantitative and qualitative variation in the genetic composition of *cas* genes. Hence, a wide range of different protein families have been defined on the basis of conserved sequences (Haft *et al.*, 2005; Makarova *et al.*, 2006).

3.2.1 Identification of *cas* genes in the *T. tenax* genome

Based on their close affiliation to CRISPR arrays, several ORFs located at TTX_1 to TTX_7 were analysed by BLASTp searches and compared to protein databases (Pfam, TIGRfam, COG and CDD). The classification of the genes was carried out according to the families and superfamilies defined by Haft and co-workers, 2005. Two regions of *cas* genes could be identified in the *T. tenax* genome, the main part was located around CRISPR loci TTX_4 to TTX_6 and a subset was found at TTX_1 (Fig. 3.6, detailed Table, Appendix A3). No *cas* genes related to previously characterised homologs could be identified next to CRISPR loci TTX_2, TTX_3 and TTX_7.

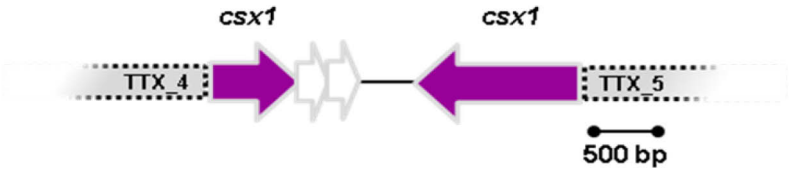
Sequence comparisons showed, that the *T. tenax* genome contained a multiplicity of *cas* genes. Thus, at least 22 genes could be assigned to one of the 45 gene families defined by comparative analysis (Haft *et al*., 2005). The core genes (*cas1-5*), encountered in the majority of prokaryotic genomes, were clustered with some archaeal *cas* genes (*csa1-5*) between TTX_5 and TTX_6. Function predictions and sequence similarities of the encoded proteins suggested, that many genes code for RNA- and DNA-modifying enzymes (Makarova *et al.*, 2006). Also, a reduced formation of the RAMP gene cluster was detected upstream of TTX_4. Interestingly, a subset of *cas* genes was identified upstream of TTX_1. Here, only the core genes *cas3* and *5* were found, but also the *cas6* gene, which is usually observed in the genomic context of the RAMP gene cluster.

a) RAMP gene cluster



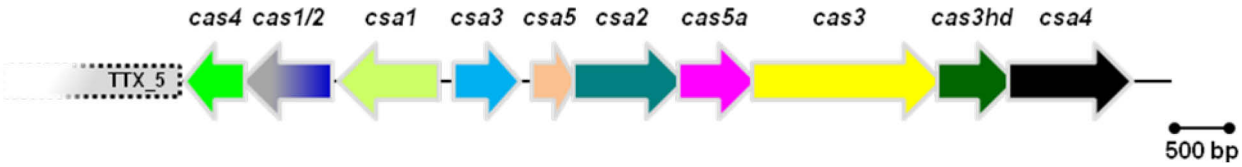
Gene-No.	Name	Gene (bp)	MW (kDa)	Putative Function
TTX_1228	Cas_Csx1	1421	53	conserved, DxTHG-motif
TTX_1229	-	381	14	hypothetical
TTX_1230	Cas_Csm1	2250	83	conserved
TTX_1231	Cas_Csm5	996	37	RAMP superfamily, RNA-binding
TTX_1232	Cas_Csm4	819	30	RAMP superfamily, RNA-binding
TTX_1233	Cas_Csm3	924	35	RAMP superfamily, RNA-binding
TTX_1234	-	426	16	hypothetical
TTX_1235	-	465	17	hypothetical
TTX_1236	-	291	11	ATP-dependent endonuclease
CRISPR TTX_4				

b) Middle section



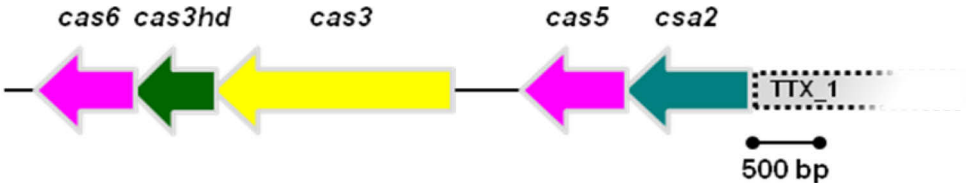
Gene-No.	Name	Gene (bp)	MW (kDa)	Putative Function
CRISPR TTX_4				
TTX_1240	Cas_Csx1	678	24	conserved, DxTHG-motif
TTX_1241	-	198	7	ATPase domain
TTX_1242	-	171	7	hypothetical
TTX_1243	Cas_Csx1	1287	46	conserved, DxTHG-motif
CRISPR TTX_5				

c) Core gene cluster



Gene-No.	Name	Gene (bp)	MW (kDa)	Putative Function
CRISPR TTX_5				
TTX_1245	Cas4	573	21	RecB-family exonuclease
TTX_1246	Cas1 / Cas2	795	29	RNA endonuclease
TTX_1248	Cas_Csa1	837	31	Cas4-like, archaeal specific
TTX_1249	Cas_Csa3	615	22	DNA-binding
TTX_1250	Csa5	393	15	coiled-coil structure
TTX_1251	Cas_Csa2	990	36	DevR_archaea, DNA-binding
TTX_1252	Cas5a	681	25	RAMP superfamily, RNA-binding
TTX_1253	Cas3	1701	63	DEAD/DEAH-box helicase
TTX_1254	Cas3_HD	681	25	HD superfamily hydrolase
TTX_1255	Csa4	1086	40	conserved, archaeal specific
CRISPR TTX_6				

d) Subset gene cluster



Gene-No.	Name	Gene (bp)	MW (kDa)	Putative Function
TTX_0232	Cas6	750	28	RAMP superfamily, RNA nuclease
TTX_0232a	Cas3_HD	572	21	HD superfamily hydrolase
TTX_0233	Cas3	1791	67	DEAD/DEAH-box helicase
TTX_0234	Cas5	783	29	RAMP superfamily, RNA-binding
TTX_0235	Cas_Csa2	936	35	DevR_archaea , DNA-binding
CRISPR TTX_1				

Fig. 3.6: Genome organisation of *cas* genes and CRISPR loci in *T. tenax*. Listed are all identified *cas* genes and selected attributes (gene and protein length, putative function). Similar colours indicate the same gene family according to Haft *et al.*, 2005; without colour: hypothetical genes, CRISPR depicted as grey boxes. ORFs analysed

by Pfam, TIGRfam, COG, CDD. **a)** RAMP gene cluster, TTX_1228-1236: *csx1* (DxTHG motif), *csm1-5* (*Mycobacterium tuberculosis* subtype). **b)** Middle section, TTX_1240-1243. **c)** Core gene cluster, TTX_1245-1255: core genes *cas1-5*, *csa1-5* (*Aeropyrum pernix* subtype). **d)** Subset gene cluster, TTX_0232-0235: core genes *cas3*/*5*/*6*

3.2.2 Transcription of *cas* genes in dependence of abiotic stress factors

In general, the CRISPR/Cas defence system is considered as a protection system of bacterial and archaeal organisms against mobile genetic elements (Barrangou *et al.*, 2007). However, some efforts have been made to show that abiotic stress factors modify the transcription levels of *cas* genes. The transcription of single *cas* genes or entire *cas* operons are shown to be regulated by environmental influences, e.g. UV-light, gamma irradiation and temperature (Williams *et al.*, 2006; Götz *et al.*, 2007; Boonyaratanakornkit *et al.*, 2007). So, the question emerged how strong is the impact of selected stress situations for the transcription of *T. tenax cas* genes.

3.2.2.1 Cell cultivation and RNA preparation

At first, an optimal cell cultivation method had to be determined to find a compromise between handling of culture volumes and RNA yield. After numerous experiments, heterotrophic growth conditions under N_2-atmosphere in a liquid minimal media volume of 4 x 500 ml were selected (s. 2.3.1). Not affected cells served as a control and therefore growth of the culture was followed for about 50 h at 86°C until the culture reached a level of 8-9 x 10^7 cells x ml^{-1} (Fig. 3.7).

For the production of stressed *T. tenax* cells the following conditions were applied: (i) Ionic stress was caused by adding an appropriate volume of a 5 M stock solution to the cell suspension and subsequently incubated for additional 3 h at 86°C until centrifugation; (ii) temperature shifts were performed by incubation of the cell culture in a water bath for 3 h at 70°C or 91°C, respectively; (iii) UV-light stress was achieved by irradiating the suspension for 30 sec or 2 min and subsequently incubated for additional 3 h at 86°C until centrifugation (s. 2.3.1.1 – 2.3.1.3). The growth conditions were equally conducted and stress parameters were applied at a similar cell density.

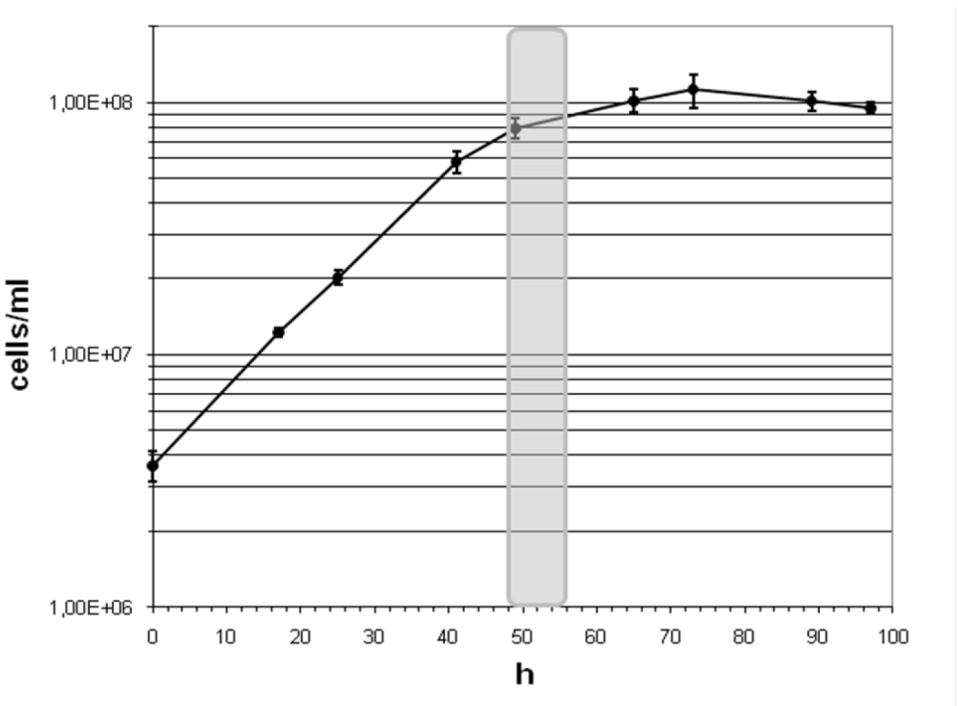


Fig. 3.7: Standard growth curve of *T. tenax*. Cells were grown heterotrophically in liquid minimal medium and density was measured by cell counting. Grey box shows the time range at the late-exponential phase (approx. 8-9 x 10^7 cells x ml^{-1}) for stress exposure and centrifugation of cells.

Total RNA was prepared from 0.15 g cell pellet using the RNeasy Mini Kit (QIAGEN) and for that purpose the instruction's manual was modified by using cell disruption with TRIzol (s. 2.5.2.2). To eliminate contamination with genomic DNA, an on-column DNase treatment was performed for 15-20 min at 30°C and the subsequent wash steps were carried out according to the manufacturer's instructions. The yield of total RNA averaged between 10-20 µg per 0.1 g cell pellet with an A_{260}/A_{280} ratio of 2.0 to 1.85. A standard Taq-Polymerase PCR assay with 500 ng total RNA as a template was performed, to control if the RNA solutions were DNA-free (s. 2.6.7.1). The 840 bp sequence of the *T. tenax* hexokinase gene (*hxk,* TTX_0060) served as a target (Fig. 3.8). In subsequent steps, only those RNA samples were used, which gave no correspondent signal in the agarose gel electrophoresis (s. 2.6.5.1). The integrity of the prepared RNA was further checked by separation of RNA on 1 % denaturing agarose gel (s. 2.5.4.1, Fig. 3.8).

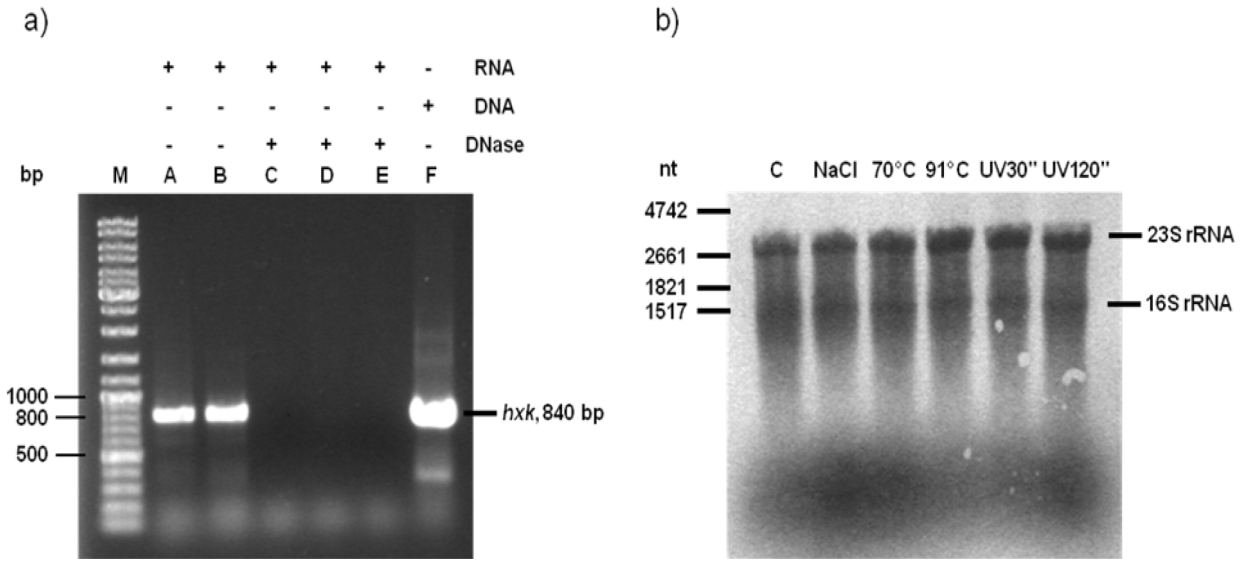


Fig. 3.8: Qualitative control of total RNA from control and stressed cultures. Total RNA preparations of *T. tenax* control and stressed cells analysed for DNA contamination in PCR assays and for integrity of immobilised methylene stained RNA. **a)** 1 % agarose gel of standard Taq-Polymerase PCR assay samples with 500 ng total RNA as template, Primer set: *hxk* for and *hxk* rev. Lane M: Gene Ruler™ DNA Ladder Mix, Lane A and C: total RNA of 50 mM NaCl stressed cells, lane B and D: 91°C stressed cells, lane E: UV-light 120" stressed cells, lane F: 80 ng genomic *T. tenax* DNA template. **b)** 2.5 µg total RNA of control (C), 50 mM NaCl (NaCl), temperature shifted (70°C, 91°C) and UV-light irradiated (UV30", UV120") cells, separated on 1 % denaturing agarose gel, marker: RNA Molecular Weight Ladder II.

3.2.2.2 Transcription under temperature and UV-light stress (RT-PCR Southern blot I)

The core *cas* genes of the CRISPR/Cas system in *T. tenax* were located between the cluster TTX_5 and TTX_6 (ORFs: TTX_1245-1255). Since the *cas1-5* core genes have been considered as essential for the molecular protection mechanisms of CRISPR systems, *cas3*, *cas4* and *csa3* were selected to study their transcription levels under different stress situations. Accordingly, Northern blot analyses were performed, with 5 µg electrophoretically separated total RNA (denaturing agarose gel, s. 2.5.4.1) and hybridised with DIG-labelled antisense RNA probes of the selected genes *cas3*, *cas4* and *csa3* (approx. length of 500 bp, s. 2.5.6). 50 ng/ml RNA antisense probes were used in hybridisation and incubated at 68°C overnight with immobilised RNA (s. 2.5.7.1). Unfortunately, in several experiments no significant signals appeared even after an exposure time up to 6 h. In control experiments, the reliability of the Northern blotting method was checked by hybridisation of 5 µg total RNA with a specific 16S rRNA probe (s. 2.5.6), DIG-labelled and handled the same

way as mentioned for the *cas* probes. This blots showed strong signals after a few seconds, which confirmed the appropriate sample preparation of the used RNA and verified the method (data not shown).

To increase the sensitivity for detecting *cas* transcripts the two-step RT-PCR Southern blot method was applied. For that, the RNA was reversely transcribed into cDNA by the M-MuLV reverse transcriptase with the help of random hexameric primer (s. 2.5.11). In the next step, this primary cDNA was amplified in a *Taq*-polymerase PCR assay with gene specific primer (s. 2.6.7.2) and lastly the secondary cDNA analysed by Southern blotting and hybridisation with gene specific RNA probes (s. 2.6.13 and 2.5.7.1). 2.5 µg of total RNA prepared from the various cultures (control and stressed cells) were applied for producing primary cDNA in a 20 µl reaction volume. In the following *Taq*-polymerase PCR amplification step, 2 µl of primary cDNA were used as a target in a 50 µl PCR reaction volume and mixed with gene specific primers for *cas4*, *cas3* and *csa3* (Tab. 3.5). The 16S rRNA served as an internal standard to ensure equal amounts of cDNA. Therefore all steps were handled in the same way. In a negative control 2.5 µg of total RNA was mixed without M-MuLV reverse transcriptase in the cDNA generating process.

Tab. 3.5: Gene specific primers used in secondary cDNA generation. *cas4*, *cas3*, *csa3* and 16S rRNA primer sequences and PCR annealing temperatures, note: primer sets for *cas4* and *cas3* produced slightly longer products as the gene length to amplify also adjacent overlapping gene areas.

Primer	Sequence (5' - 3')	T_m	T_m PCR	Length (bp)
cas4 for	CCAGACCTCCCGCGGCATATGTC	57°C	62°C	617
cas4 rev	GGAGTTTAGGCCGGATCCGTGAGG	57°C		
cas3 for	GGCTGTGTTCCCTGAGGCATATGGTT	64°C	63°C	1740
cas3 rev	CGGAAGTAGGCGCAGGAATTCATCG	62°C		
csa3 for	GCAGGTGGCTTCATATGAGGATGGGC	66°C	62°C	614
csa3 rev	ATTAAGGATCCCTGTCTACAAAGATCCAGCTC	63°C		
16S for	TTCCGGTTGATCCTGCCGGA	55°C	52°C	1509
16S rev	GGTTACCTTGTTACGACTT	52°C		

To identify the amplified PCR products by hybridisation with the specific antisense RNA probes, 20 µl of the secondary cDNA were electrophoretically separated on a 1% agarose gel and afterwards the gel incubated in a denaturation and neutralisation solution to make the separated DNA molecules susceptible for hybridisation (s. 2.6.13). After blotting the cDNA on a positively charged nylon membrane, the blot was incubated overnight at 52°C with the DIG-labelled RNA probes (s. 2.5.7.1).
The results of three independent experiments indicated, that the transcription levels of the *cas3* and *cas4* genes were affected by the given stress parameter (Fig. 3.9). The *cas3* gene showed a more than threefold increased and the *cas4* gene a twofold increased transcript level in cells treated 2 min with UV-light in comparison to the control cells. Furthermore, the mRNA levels of the *cas3* genes were decreased eight-fold in temperature shifted cells and in cells treated 30 sec with UV-light. The transcript levels of the *csa3* gene and the 16S rRNA were not significantly affected and the negative control showed no detectable signal after a comparable exposure time of 15 min. Slight differences in signals of the 16S control were factored in all other values.

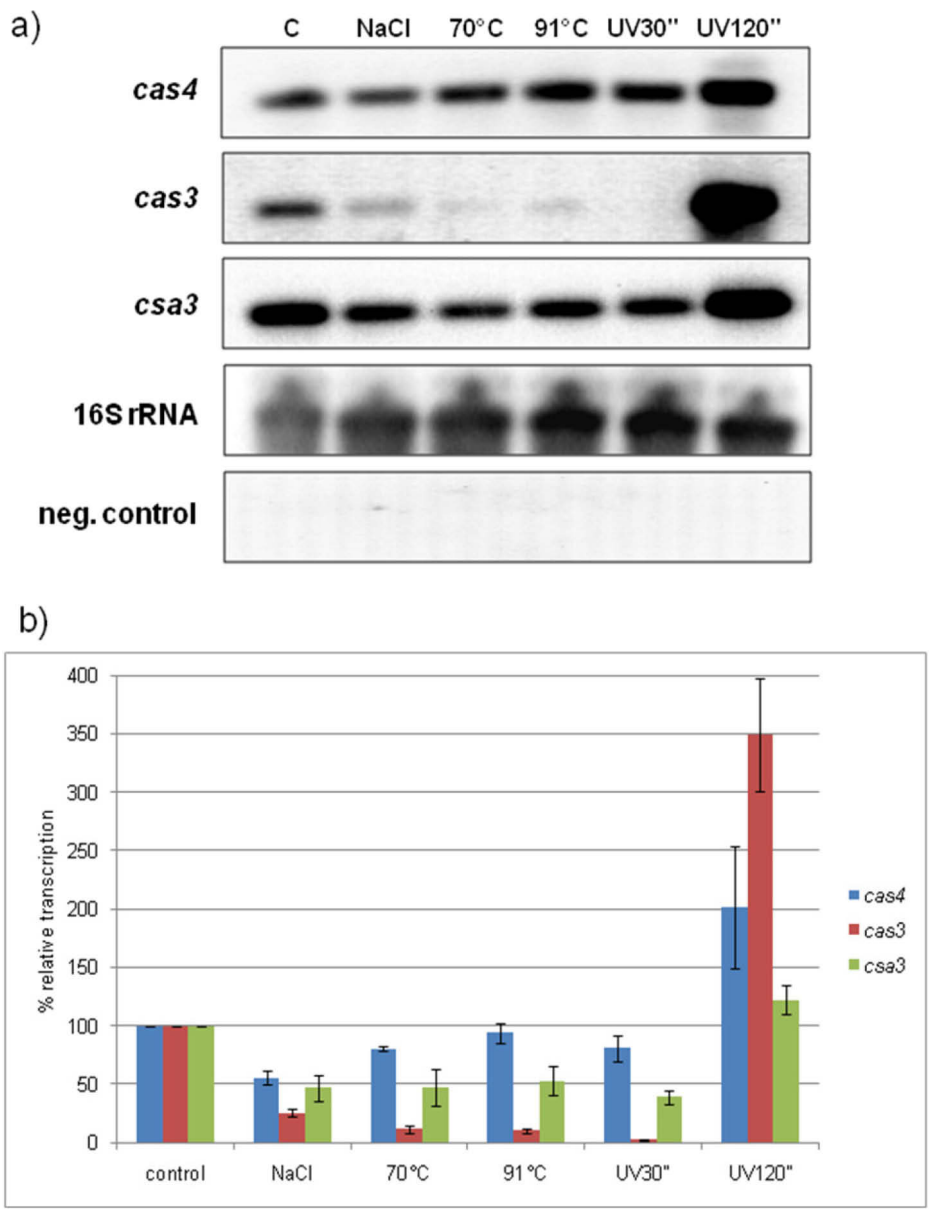


Fig. 3.9: RT-PCR Southern blots I and data interpretation of signals. a) mRNA levels of *cas4*, *cas3* and *csa3* in control cells (C) and stressed cells (NaCl: 50 mM NaCl, 70°C/91°C: temperature shift, UV30''/UV120'': UV-light) analysed after electrophoretic separation of amplified cDNA in a Southern blot. 16S rRNA internal standard, negative control without RT-step and amplified with primer set *cas3*. **b)** Data quantification of signals with ImageJ of three independent experiments. Transcript levels of control cells set to 100 % and compared to transcript levels of stressed cells. Differences in internal standard 16S were factored in all values. Black bars show standard deviation.

3.2.2.3 Transcription under ionic stress (RT-PCR Southern blot II)

The influence of the NaCl concentration on the growing behaviour of *T. tenax* was studied (e.g. generation time, max. cell counts) to verify the stress potential of ionic strength on the organism. For that, batch cultures of *T. tenax* in the presence of 50,

100 and 150 mM NaCl were followed over 50 h and compared to control cells without NaCl. The medium was inoculated with 4 % of a preparatory culture (1 x 10^8 cells x ml^{-1}) and incubated at 86°C for 15 h (s. 2.3.1). At that time point all cultures had a similar cell count (1 x 10^7 cells x ml^{-1}) and were then incubated with or without the respective NaCl concentrations. The growth rates and maximal cell counts (2 x 10^7 cells x ml^{-1}) showed that 100 and 150 mM NaCl affected the viability of the cells in comparison to standard grown cells. 50 mM NaCl reduced slightly the growth of the culture with a maximal cell count of 8 x 10^7 cells x ml^{-1} (Fig. 3.10).

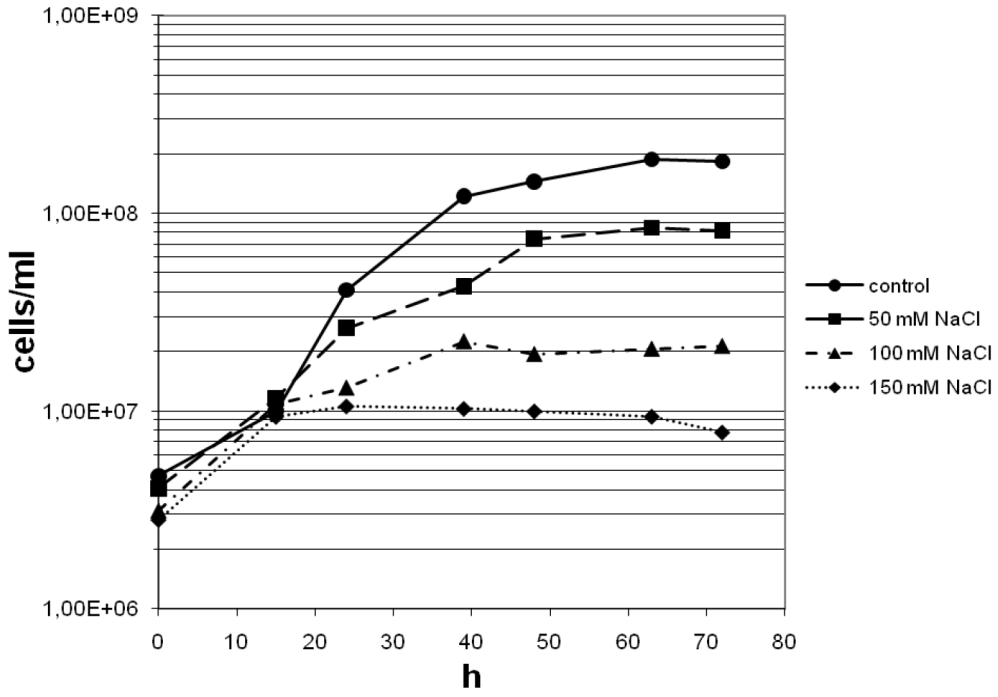


Fig. 3.10: Batch cultures of *T. tenax* with and without NaCl. Cell culture incubated at the same time point (15 h) with 0, 50, 100 and 150 mM NaCl.

These results indicate that the previously selected ionic stress conditions were too gentle. In the first transcription experiments, the ionic stress was only increased by addition of 50 mM NaCl for 3 h. The results showed no significant effect of NaCl on the transcription levels of the *cas* genes (Fig. 3.9).

For the subsequent RT-PCR Southern blot experiments it was decided to use the above mentioned concentrations of NaCl. Moreover, the stress incubation time was extended from 3 to 6 h to facilitate probable differences in transcript levels. The *T. tenax* cells were grown heterotrophically until late exponential phase, the respective

NaCl concentration was increased to 50, 100 and 150 mM by injecting the respective volume of a 5 M NaCl solution and further incubated for 6 h at 86°C. From the cultures approx. 7.5-15 µg per 0.1 g cell pellet could be obtained (A_{260}/A_{280} ratio: 1.9) and RNA integrity was further checked (s. 2.6.5.1). 1.5 µg total RNA was applied in the reverse transcription assay (s. 2.5.11), 2 µl of primary cDNA was again used as a template in *Taq*-polymerase PCR assay with the primer sets for *cas4*, *cas3* and *csa3* (s. 2.6.7.2, Tab. 3.5) and subsequently cDNA was identified by Southern blots (s. 2.6.13).

The results of two independent experiments showed, that the transcription level of the *cas3* gene was clearly affected by the ionic strength (Fig. 3.11). The *cas3* gene showed at 100 mM NaCl a more than tenfold increased and at 150 mM NaCl still a fivefold increased mRNA level in comparison to the control. However, at 50 mM NaCl the transcript level was sevenfold decreased. The *csa3* mRNA levels were slightly affected, only twofold increased signal intensity was observed at 50 mM NaCl. The *cas4* and the 16S rRNA transcripts were nearly unaffected. Taken together, RT-PCR Southern blot experiments clearly show that especially the mRNA level of the *cas3* gene was strongly increased under higher ionic strength and UV-light dose rates.

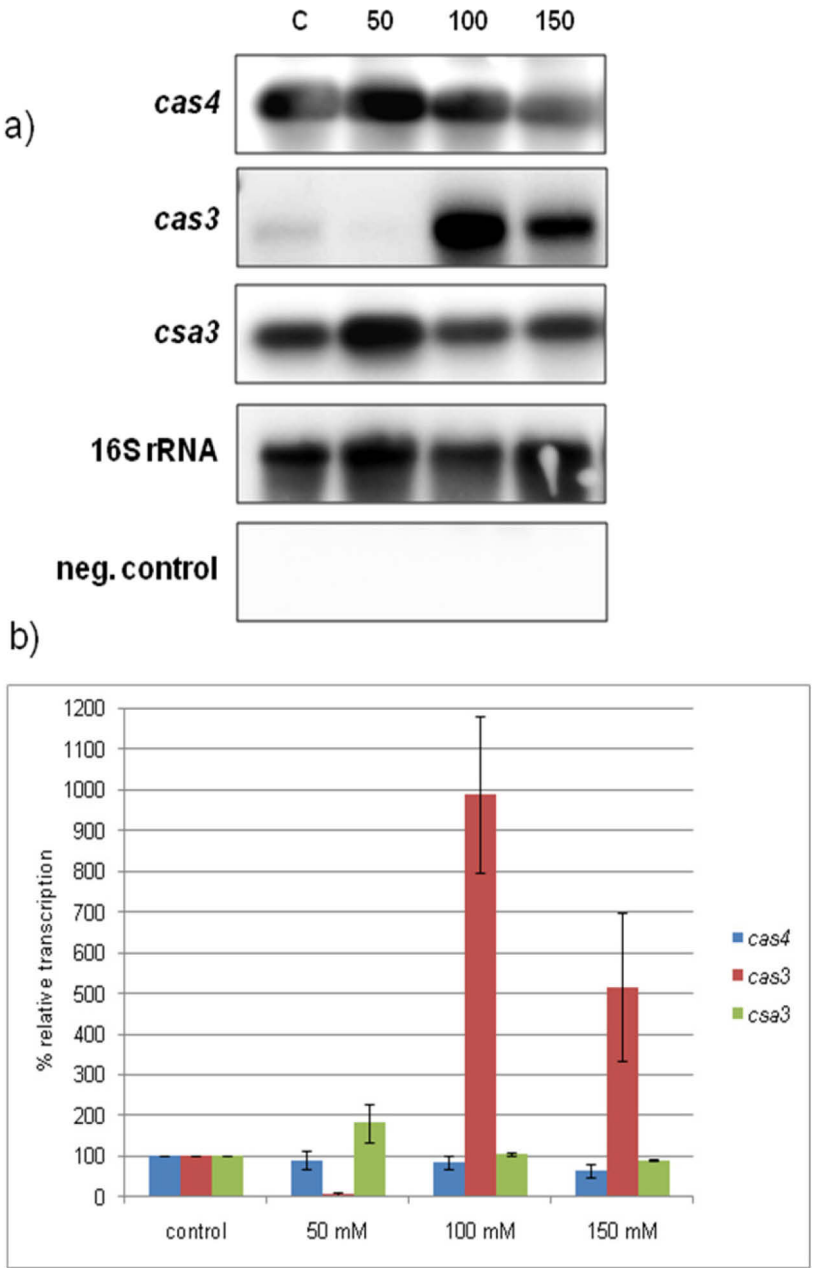


Fig. 3.11: RT-PCR Southern blots II and data interpretation of signals. a) Transcript levels of *cas4*, *cas3* and *csa3* in control cells (C) and in cells grown at 50, 100 and 150 mM NaCl determined by electrophoretic separation of amplified cDNA and Southern blot analyses. 16S rRNA served as internal standard, negative control without RT-step and amplified with primer set *cas3*. **b)** Data quantification of signals with ImageJ of two independent experiments. Control cells were set to 100 % and compared to mRNA levels of ionic stressed cells. Differences in internal standard 16S were factored in all values. Black bars show standard deviation.

3.2.3 Detection of polycistronic transcripts of core *cas* genes

An operon is defined as a structural and functional unit of a genome containing a cluster of genes under the control of a regulatory signal or promoter. The structural genes of an operon are transcribed into one polycistronic mRNA, i.e. an mRNA

molecule that codes for more than one protein. Upstream of the structural genes a promoter sequence is located, which provides a RNA polymerase binding site to initiate the transcription. Close to the promoter lies a section of DNA called an operator. Because of the clustering, gene orientations and overlapping start and stop codons of genes, the ORFs of the core *cas* gene cluster were assumed to form two operon structures (coding range of operon *casa1*: *cas4*, cas1/2 and *csa1*, TTX_1245-48; coding range of operon *casa2*: *csa5*, *csa2*, *cas5a*, *cas3*, *cas3hd* and *csa4*, TTX_1250-55). To define the polycistronic mRNA of both *cas* operons, 3 µg prepared total RNA derived from control cells (s. 2.5.2.2) were reversely transcribed into cDNA (s. 2.5.11) and 1 µl was used as template in a standard *Taq*-polymerase PCR reaction (s. 2.6.7.2). The primers were generated in this way that the PCR transcripts covered areas of overlapping gene sequences (Tab. 3.6) and the amplified PCR products were then analysed by agarose gel electrophoresis (s. 2.6.5.1).

Tab. 3.6: PCR primers for detection of polycistronic mRNAs *casa1* and *casa2*. Primers and PCR annealing temperatures for detection of overlapping mRNA regions

Primer	Sequence (5' - 3')	T_m	T_m PCR	Length (bp)
casa1a for	CCCAAGACGGGATCCTACTACAGCATATG	58°C	59°C	938
casa1a rev	CGTATTAGGATCCGGTTAATGCAGAGCG	56°C		
casa1b for	CATCAAAACCATATGTTGACACTCCTGG	53°C	62°C	2315
casa1b rev	GGAGTTTAGGCCGGATCCGTGAGG	57°C		
casa2a for	ATGGAGTCTGTACAACTCAAGAAG	49°C	59°C	1382
casa2a rev	GCATCACCTCCTAGCGGGTTGCC	57°C		
casa2b for	TGTATTCTCTTCCATATGGAGTCT	47°C	54°C	1424
casa2b rev	GGCGCCCTGGCCTGGATCC	57°C		
casa2c for	CCTGAGGGAGATGGTTGGTAG	51°C	58°C	1730
casa2c rev	CGGAAGTAGGCGCAGGAATTCATCG	56°C		
casa2d for	AGTGAGCGAGCTGTGCATATGAC	52°C	54°C	1102
casa2d rev	AGCCTATCTTTTTAAGGGCT	43°C		
casa2e for	AGGTGATGCACTACTACTTGA	45°C	59°C	723
casa2e rev	CGCCAGGCTCGGATCCCTCACC	59°C		
casa2f for	AAGGGGGCGGAGGAGCATATGAG	56°C	58°C	726
casa2f rev	GCGTGATCATCTCCTGGATCCACTG	56°C		
casa2g for	AAGGGGGCGGAGGAGCATATGAG	56°C	57°C	1817
casa2g rev	CGTGGGGCAGGATCCGAATTT	51°C		

The results proved that both genomic units are organised as an operon, since every overlapping part of the polycistronic transcripts could be detected by a specific PCR transcript (Fig. 3.12). Three primer sets (*casa1b*, *casa2b*, *casa2g*) served as a control, because the covered sequences crossed the border of the expected polycistronic mRNA transcripts. Although *casa1* showed a polycistronic transcript, it should be noted that a gap of 89 bp is located between the genes *csa1* and *cas1*/*2*. But obviously this is not influencing the complete transcription of the operon.

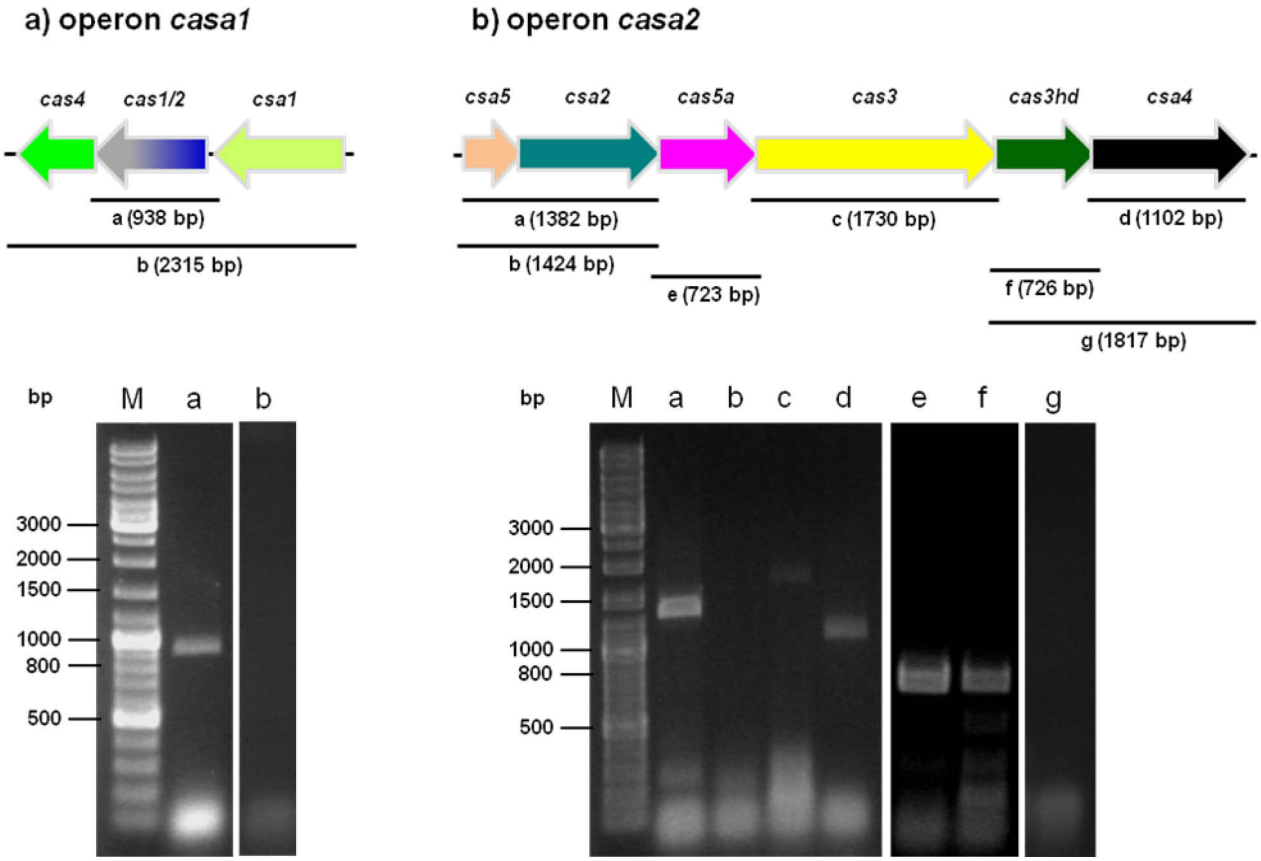


Fig. 3.12: Genome organisation and detection of polycistronic mRNAs *casa1* and *casa2*. PCR products of amplified cDNA separated on 1 % agarose gels, letters a-g in the figure correlate to primer sets in Table 3.6, M: GeneRuler™ DNA ladder mix. **a)** *casa1* operon, TTX_1245-48. Primer set for b served as a control, since binding occurred outside the polycistronic transcript of *casa1.* **b)** *casa2* operon, TTX_1250-55. Primer sets for b and g served as a control, since binding occurred outside the polycistronic transcript of *casa2.*

3.3 Functional analysis of the *T. tenax* DNA-binding protein Csa3

3.3.1 Genomic context and sequence information of *csa3*

In the classification of *cas* genes, the *csa3* gene (*cas* gene specific for *Aeropyrum pernix* subtype, Haft *et al.*, 2005) encodes a putative DNA-binding protein and homologs are identified in different archaeal organisms. To verify the putative DNA-binding function of Csa3, the protein sequence was analysed with different secondary structure prediction tools for the identification of DNA-binding motifs. Helix-turn-helix (HTH) prediction using the NPS@ (Network Protein Sequence Analysis; Combet *et al.*, 2000) suggested that the Csa3 protein contained a HTH DNA-binding motif (sequence at position 145-166: VGIDDVAQLVGRDSTTVARYLK; Appendix A4) with an approx. 25 % probability. Also the GYM 2.0 prediction software suggested a HTH-motif at the same position with 58 % probability. Other DNA-binding motifs, such as the zinc finger (C_2H_2 ZNF database) or the leucine zipper (2Zip server) were not detected. In all likelihood, the Csa3 protein shall contain a HTH-motif and therefore possess the ability for DNA-binding. Furthermore, Csa3 showed not only significant similarity to other archaeal Csa3 proteins, but also to the TrmB protein (sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter) of *P. aerophilum* (E-value: 2e-29).

In *T. tenax*, the putative *csa3* gene was located between the two gene clusters of *casa1* and *casa2*. Transcript analyses of *casa1* and *casa2* showed that both genomic cluster formed operon structures, since polycistronic mRNAs could be detected (Fig. 3.12). Up- and downstream of the *csa3* gene the putative promoter regions of *casa1* and *casa2* were localised and characterised by well conserved sequences (Soppa, 1999): the *Archaea* specific BRE-site (RNWAAW) and TATA-box (YTTTTAAA). The promoter elements for *casa1* and *casa2* could be defined in comparison to known promoters from *T. tenax* or other crenarchaeal organisms. For *csa3* no conserved sequence elements could be found in the region -10 to -35 (position 0 accords to translation start codons ATG or GTG). Only at positions -45 to -60 acceptable recognition elements were localised. Thereby, the putative TATA-box element would be simultaneously responsible for the transcription of *csa3* and *casa1* (Fig. 3.13).

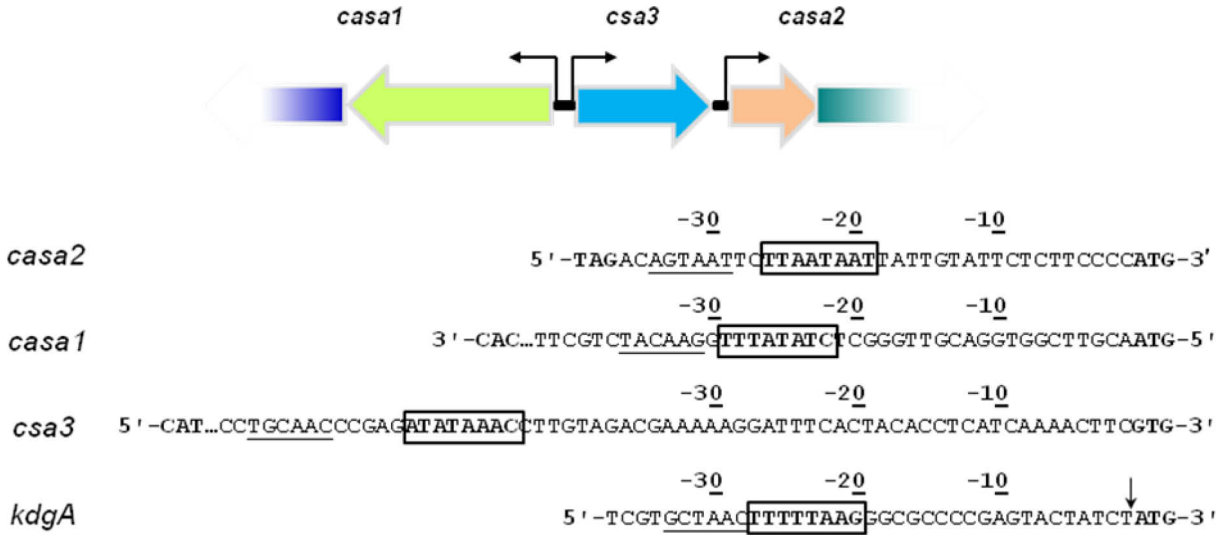


Fig. 3.13: Putative BRE-site and TATA-box elements of promoter regions *casa1*, *casa2* and *csa3*. Promoter elements defined in comparison to archaeal consensus sequences for the BRE-site (consensus: RNWAAW) and TATA-box (consensus: YTTTTAAA). Position 0 accords to translation start codons ATG or GTG. Putative transcription start sites (arrows), BRE-sites (underlined) and TATA-boxes (black box) marked. The transcript start of *kdgA* (2-keto-3-deoxy-(phospho)gluconate aldolase) was determined by primer extension analyses (Ahmed *et al*., 2005) and served here as an example of *T. tenax* (arrow).

To verify the DNA-binding function of Csa3 experimentally, electrophoretic mobility shift assays (EMSAs) were performed. Since Csa3 is supposed to have a function as transcriptional regulator, the proximal promoter regions *casa1* and *casa2* were selected as DNA targets (Tab. 3.7). The respective promoter sequences were amplified by *Taq*-polymerase PCR assays (s. 2.6.7.1), separated on an 8 % non-denaturing PAGE (s. 2.6.5.2) and the PCR products extracted by electroelution from the gel matrix (s. 2.6.6.3). The purified DNA fragments were 3'-end DIG-labelled and precipitated (s. 2.6.11.1).

Tab. 3.7: Oligonucleotides for amplification of the promoter spanning regions *casa1* and *casa2*. Primer sets used in *Taq*-polymerase PCR assays. Amplified promoter regions slightly longer as depicted in Fig. 3.13.

Name	Sequence (5' - 3')	T_m	T_m PCR	Probe length
Pro-*casa1*-for	AACCCCACCGTGATGCCCAT	51°C	60°C	150 bp
Pro-*casa1*-rev	GGTGACCTTGGCCCTCCTCA	53°C		
Pro-*casa2*-for	TTCGTCGACCCGCTCCTCACCA	55°C	60°C	145 bp
Pro-*casa2*-rev	CGGCCCGCTCCACTAGCTCT	55°C		

3.3.2 Cloning, heterologous expression and purification of *csa3* in *E. coli*

For the experimental evidence of Csa3 DNA-binding capacity, the respective *csa3* gene was cloned and heterologously expressed in *E. coli*. The *csa3* gene (615 bp) was amplified by PCR mutagenesis with *Pfu* DNA-polymerase using 100 ng genomic *T. tenax* DNA as a template and the primer sets *csa3* for and *csa3* rev (Tab. 3.8). For recombinant expression using the pET system, the amplified gene was cloned into pET24a and the sequence was checked by automated dideoxy sequencing (s. 2.6.10). Expression was performed in *E. coli* Rosetta(DE3) transformed with the recombinant vector pET24a-*csa3* (s. 2.7.1).

Tab. 3.8: Primer sets for amplification of *csa3*. Primer sets used in *Pfu*-polymerase PCR mutagenesis assays. Restriction sites underlined. PCR settings: elongation 1:30 min at 72°C.

Name	Sequence (5' - 3')	Restriction	T_m	T_m PCR
csa3 for	GCAGGTGGCTTCATATGAGGATGGGC	*Nde*I	58°C	62°C
csa3 rev	ATTAAGGATCCCTGTCTACAAAGATCCAGCTC	*Bam*HI	57°C	

Purification of the recombinant expressed Csa3 protein revealed some unexpected characteristics. At the beginning, the protein was extracted by cell disruption, enriched in the heat precipitation step at 80°C and purified by cationic exchange chromatography with heparin-sepharose (s. 2.7.2.3). In a SDS-PAGE the corresponding fractions that contained Csa3 (21 kDa) were identified and pooled. Surprisingly, the protein solution showed abnormal absorption values at 280 nm in spectrometric measurements. Subsequent analyses by agarose gel electrophoresis revealed that the protein binds high amounts of DNA, which explained the unusual absorption at 280 nm. The strong DNA binding of Csa3 resisted the different steps of the purification process (Fig. 3.14). Binding experiments of the purified protein with the promoter region *casa1* and *casa2* as DNA target, showed no retardation in the EMSAs (data not shown). Presumably, the DNA contamination caused a blocking of specific DNA-binding sites of Csa3.

Therefore, a protein purification protocol was established for removing of DNA contamination by polyethylenimine (PEI), which can be used as precipitant for the negatively charged nucleic acids. Consequently, the heat precipitated Csa3 protein solution was treated with 0.5 % PEI pH 8 by adding dropwise PEI to the chilled and stirring solution. The PEI-DNA complexes could be removed by centrifugation and

additional ammonium sulphate (AS, 80 %) precipitation. The Csa3 protein was again purified by cationic exchange chromatography with heparin-sepharose (s. 2.7.2.3). The subsequent analysis on an agarose gel showed that the DNA contaminations were efficiently removed by this method (Fig. 3.14). From 1 g recombinant *E. coli* cells about 2.5 mg/ml purified Csa3 protein (~ 95 % purity) was obtained (Fig. 3.15).

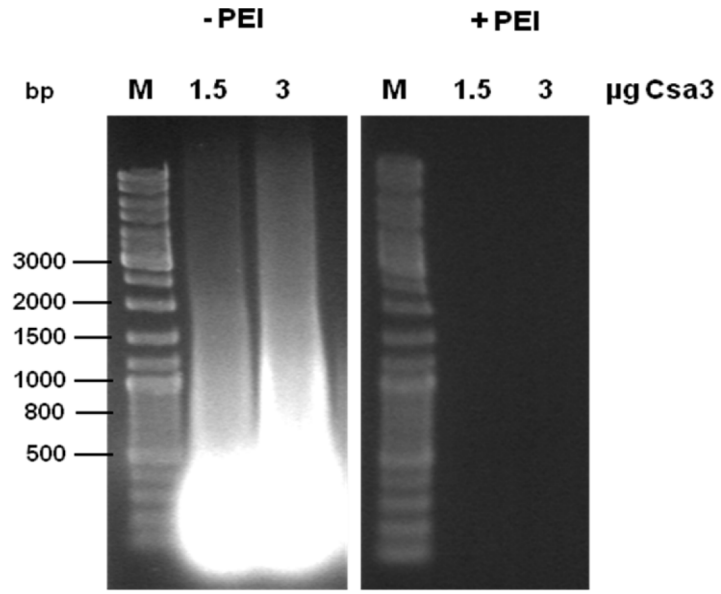


Fig. 3.14: Agarose gel electrophoresis of Csa3. Comparison of protein preparations treated with and without PEI. 1.5 and 3 µg Csa3 purified by heparin-sepharose was separated on 1 % agarose gel. M: GeneRuler™ Ladder Mix.

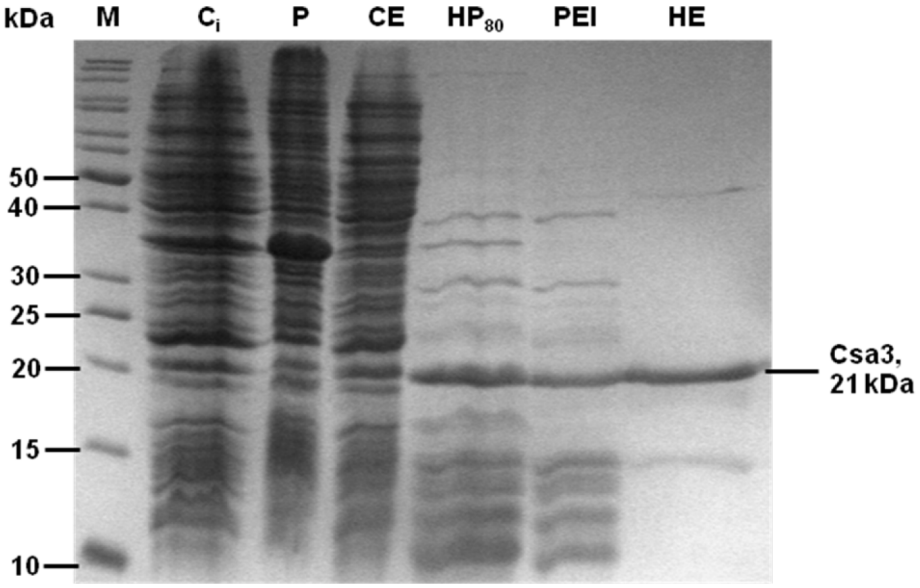


Fig. 3.15: Purification of the recombinant Csa3. 15 % SDS-PAGE, M: Unstained Protein Ladder, C_i: lysed *E. coli* cells 3 h after IPTG induction, P: pellet fraction after centrifugation of cell extracts, CE: crude extract, supernatant after centrifugation, HP_{80}: heat precipitated crude extract at 80°C, PEI: heat precipitated extract after PEI and AS precipitation, HE: pooled and concentrated fractions after heparin-sepharose.

3.3.3 DNA-binding studies using EMSAs

The generation of DNA probes was performed by PCR amplification and 3'-end DIG labelling using terminal transferase (s. 2.6.11.1). The PCR products contained the putative DNA-binding sites of the promoter regions *casa1* and *casa2*. 40 ng of the DIG-labelled probes and 50 to 300 ng of the purified and DNA-free recombinant Csa3 protein were incubated for 10 min at 50°C in binding buffer (s. 2.6.14). The DNA-protein complexes were separated from the free DNA by 1 % agarose gel electrophoresis followed by the transfer to a positively charged nylon-membrane. Finally, the immunological reaction was carried out for the detection of DNA-protein complexes (s. 2.6.15 and Fig. 3.16). The EMSA showed that the protein binds the promoter region of *casa1* and *casa2* with slight preference for *casa1* already at 5 ng/µl, as indicated by stronger reduction of free DNA. But obviously only faint shifts and a smearing of signals were observed, accounting for an instable or unspecific binding of Csa3 to the promoter DNA.

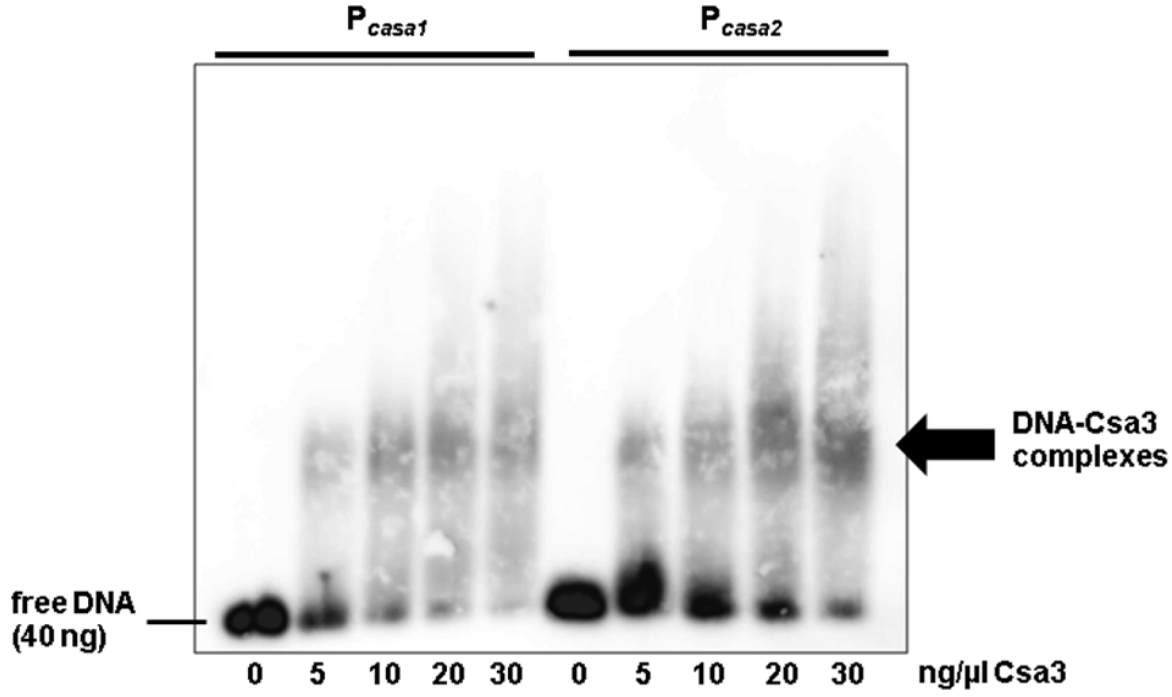


Fig. 3.16: EMSA with Csa3 and the promoter spanning DNA regions *casa1* and *casa2*. Different quantities of purified Csa3 (0-30 ng/µl) incubated with 40 ng DIG-labelled probes (promoter spanning regions: P_{casa1} and P_{casa2}) at 50°C. Faint shifts appeared, due to DNA-protein binding at 5-30 ng/µl Csa3. Putative DNA-protein complexes are marked by an arrow.

To prove specificity of the DNA-Csa3 interaction, competition experiments with 100 ng synthetic Poly dI-dC DNA were performed. However, the results indicate that the

DNA-Csa3 interaction is rather unspecific. Also higher used protein concentrations (up to 80 ng/µl Csa3) showed the same effect. Taken together, the putative transcriptional regulator Csa3 showed DNA-binding activities. But the purification and EMSA experiments suggested that the DNA-binding of the protein to the promoter regions of *casa1* and *casa2* was unspecific under the chosen conditions.

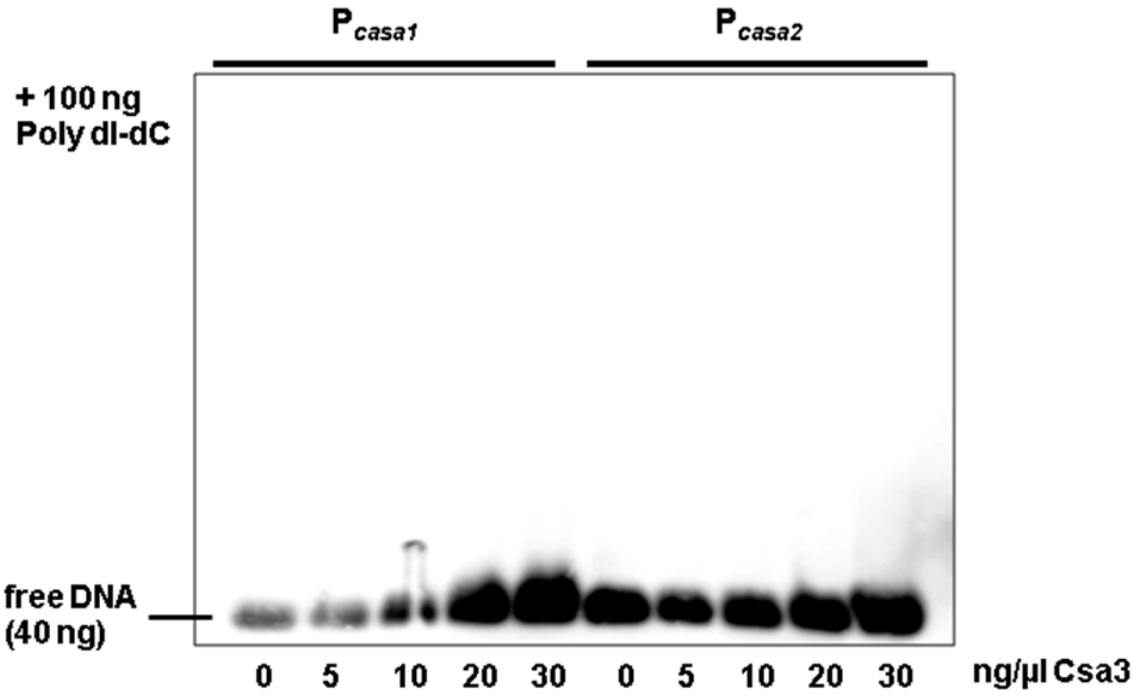


Fig. 3.17: **EMSA with Csa3 using *casa1* and *casa2* promoter region in presence of Poly dI-dC as competitor DNA.** Different quantities of purified Csa3 (0-30 ng/µl) incubated with 40 ng DIG-labelled probes (promoter spanning regions: P_{casa1} and P_{casa2}) at 50°C in presence of 100 ng competitor DNA (Poly dI-dC).

3.4 Analyses of the archaeal CRISPR-associated complex 1 (CasA1)

3.4.1 Genomic context and sequence information of *casa1*

As deduced from gene arrangement with overlaps of coding regions and from RT-PCR Southern blot analyses, which clearly indicated a co-transcription of *cas4*, *cas1*/*2* and *csa1* on the one hand, and *csa5*, *csa2*, *cas5a*, *cas3*, *cas3HD* and *csa4* on the other hand, a selective interaction between the encoded proteins has to be assumed. As applicable for Cas proteins in general, the knowledge about the functional impact of the proteins encoded by the *cas* operon *casa1* and their homologs is rather low. Concerning the proteins encoded by the *casa1* operon: Cas4 encoded by the first gene of operon, represents a member of a protein family (Cas4 family), which could be identified in nearly all archaeal genomes and in a low number of *Bacteria*, e.g. *Thermotoga neapolitana* and *Desulfovibrio vulgaris*. The sequence of these proteins contains a C-terminal motif of four conserved cysteine residues and resembled the RecB-family of exonucleases. The Csa1 protein is similar to the Cas4 family and specific for crenarchaeal organisms. A unique feature of the *T. tenax* genome is the obvious fusion between *cas1* and *cas2* suggesting a close interaction between the encoded proteins. Cas1 has been predicted as an α-helical nuclease and Cas2 a metal-dependent endoribonuclease with a ferredoxin-like fold. These two proteins are the only Cas proteins conserved in nearly all analysed genomes.

3.4.2 Cloning and heterologous expression of the genes of *casa1* in *E. coli*

To understand functional impact and molecular interaction of the proteins encoded by the operon *casa1*, efforts have been made to express these proteins heterologously in *E. coli* for detailed investigations. Generally, all sequences of recombinant genes were checked by automated dideoxy sequencing (s. 2.6.10).

3.4.2.1 Cloning and heterologous expression of Cas4 and Csa1

The genes *cas4* and *csa1* were amplified by PCR mutagenesis with *Pfu* DNA-polymerase using 100 ng genomic *T. tenax* DNA as a template and the respective primer sets (primer: Tab. 3.9, PCR settings: Tab. 3.10). Additionally, the genes were cloned in frame with a C-terminal His-tag (6x histidine). Therefore, reverse primers containing the six codons (GTG) coding for histidine were constructed.

Tab. 3.9: Primer sets for amplification of *cas4* and *csa1*. Primer sets used in *Pfu*-polymerase PCR mutagenesis assays. Restriction sites (rs) underlined, GTG-codons for His-tag marked in black

Name	Sequence (5' - 3')	Rs	T_m
cas4 for	CCAGACCTCCCGCGGCATATGTC	*Nde*I	57°C
cas4 rev	GGAGTTTAGGCCGGATCCGTGAGG	*Bam*HI	57°C
cas4 his rev	GCCGGCGCCGGGATCCGCTAGTGGTGGTGGTGGTGGGATTTTTTCC	*Bam*HI	72°C
csa1 for	CATCAAAACCATATGTTGACACTCCTGG	*Nde*I	53°C
csa1 rev	TGACGCCGTATGGATCCACTACGA	*Bam*HI	54°C
csa1 his rev	GACGCCGTATGGATCCACTAGTGGTGGTGGTGGTGGCGACCTGCAT	*Bam*HI	71°C

Tab. 3.10: PCR settings for amplification of *cas4* and *csa1*. PCR settings for *Pfu*-polymerase PCR mutagenesis assays (s. 2.6.7.3), elongation at 72°C.

Product	Primer sets	T_m PCR	Elongation	Length, bp
cas4	*cas4* for / *cas4* rev	62°C	1:30 min	573
cas4 his	*cas4* for / *cas4* his rev	55°C	1:30 min	591
csa1	*csa1* for / *csa1* rev	59°C	2:00 min	837
csa1 his	*csa1* for / *csa1* his rev	53°C	2:00 min	855

For expression using the pET system, the amplified gene was cloned into pET24a and the recombinant vectors (pET24a-*cas4*, pET24a-*cas4* his, pET24a-*csa1* and pET24a-*csa1* his) were used for transformation of *E. coli* Rosetta(DE3) (s. 2.7.1).

For the purification of Cas4 and Csa1, the proteins were extracted by cell disruption and enriched in heat precipitation at 70-90°C (s. 2.7.2.1). No protein could be detected in the soluble crude extract or in the supernatant of heat precipitated extracts, suggesting that the recombinant protein was expressed in inclusion bodies, i.e. protein aggregates, which precipitate and can be separated by centrifugation (Fig. 3.18). The assumption was confirmed by purification of heat precipitated His-tagged proteins with Ni-NTA agarose (s. 2.7.2.7). No protein could be detected in the fractions after elution with 250 mM imidazole on a SDS-PAGE at the respective molecular weight size (Cas4: 21 kDa, Csa1: 31 kDa).

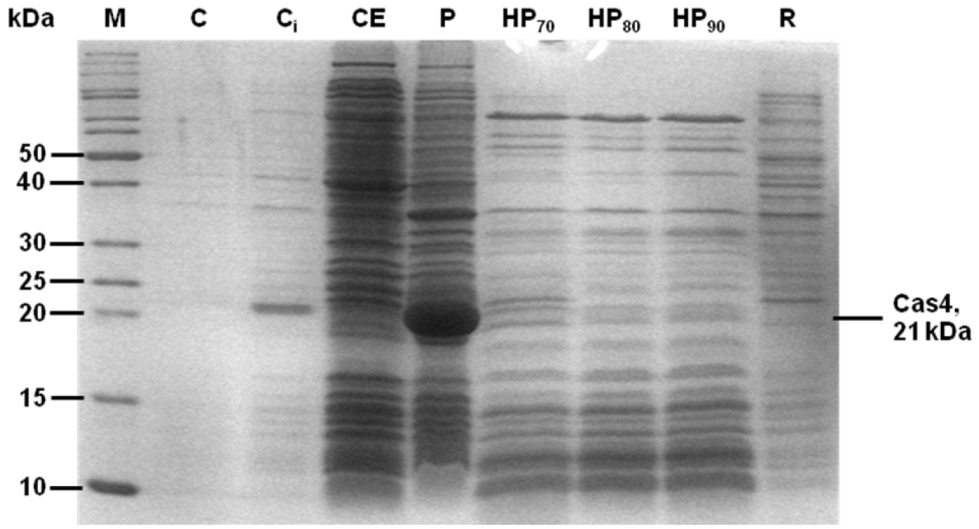


Fig. 3.18: Purification of the recombinant Cas4. 15 % SDS-PAGE, M: Unstained Protein Ladder; C: Lysed *E. coli* cells before IPTG induction; C_i: Lysed *E. coli* cells after 3 h IPTG induction; CE: crude extract, soluble supernatant after centrifugation of cell extracts; P: pellet fraction after centrifugation of cell extracts; $HP_{70, 80, 90}$: heat precipitated and centrifuged crude extract at 70, 80 and 90°C, respectively; R: *E. coli* Rosetta strain without pET24a plasmid. Insoluble Cas4 protein detected at 21 kDa in pellet fraction.

Tab. 3.11: Expression conditions modified for recombinant expression of Cas4. To avoid formation of inclusion bodies, standard *E. coli* cultivation (s. 2.3.2) and different modifications listed; protein expression analysed by SDS-PAGE.

E. coli strain	Medium	Temperature	Induction of expression / modifications
Rosetta	LB	37°C	1 mM IPTG, 3 h
Rosetta	LB	37°C	0.5 mM IPTG, 3 h
BL21	LB	37°C	1 mM IPTG, 3 h
pLysS	LB	37°C	1 mM IPTG, 3 h
RIL	LB	37°C	1 mM IPTG, 3 h
Rosetta	LB	37°C – 42°C	1mM IPTG, heat shock: upshift 42°C, 3 h
Rosetta	LB	37°C – 25°C	1mM IPTG, cold shock: downshift 25°C, 3 h
Rosetta	LB	25°C	1 mM IPTG, 3 h
Rosetta	LB	37°C	1 mM IPTG, anaerobic cultivation, o/n
Rosetta	LB	37°C – 20°C	0.1 mM IPTG, downshift 20°C, o/n
Rosetta	LB	37°C	0.1 mM IPTG, + 20 mM Tris/HCl pH 7, 3 h
Rosetta	LB	37°C	0.1 mM IPTG, + 1 % glucose, 3 h
Rosetta	LB	37°C	0.1 mM IPTG, + 300 mM sorbitol, 3 h
Rosetta	LB	37°C	0.1 mM IPTG, + 300 mM sucrose, 3 h
Rosetta	LB	37°C	0.1 mM IPTG, + 1-3 % ethanol, 3 h
Rosetta	2 x YT	37°C	0.1 mM IPTG, 3 h
Rosetta	NZA	37°C	0.1 mM IPTG, 3 h

To avoid aggregation of proteins during recombinant expression, growth conditions were varied in a wide range, e.g. cloning of the vector in different host strains or different modifications and media during expression (Kube *et al.*, 2006; Demir and Dincturk, 2006; Georgiou and Valax, 1996). Tab. 3.11 gives an overview of the modifications for expression of Cas4. But none of the chosen conditions were successful. These results suggested that the single proteins were unable to form stable properly folded structures during the expression in *E. coli*. Therefore, the aggregated proteins were purified and efforts were made to solubilise and refold the proteins (s. 2.7.2.5).

3.4.2.2 Cloning and heterologous expression of Cas1/2 and CasA1

To express the fusion protein Cas1/2, the gene (795 bp) was amplified by PCR mutagenesis for the insertion of restriction sites with *Pfu* DNA-polymerase using the primer sets *cas1*/*2* for and *cas1*/*2* rev (Tab. 3.12). For recombinant expression the amplified gene was cloned into pET15b. Additionally, the entire *casa1* operon was cloned via overlap extension PCR (s. 2.6.7.4) to remove an internal *Nde*I restriction site and facilitate ligation with vector pET24a. Two PCR amplified fragments (Tab. 3.12: primer sets: 1/2, 1456 bp and 3/4, 844 bp), overlapping 23 bp were hybridised at 56°C for 3 min to generate the template for subsequent PCR amplification with primer set 1/4. The recombinant vectors pET15b-*cas1*/*2* and pET24a-*casa1* were used for transformation of *E. coli* Rosetta(DE3) (s. 2.7.1).

Tab. 3.12: Primer sets for amplification of *cas1*/*2* and operon *casa1*. Primer sets used in *Pfu*-polymerase PCR mutagenesis assays (s. 2.6.7.3). Restriction sites (rs) underlined.

Name	Sequence (5' - 3')	Rs	T_m	T_m PCR
cas1/*2* for	GGTAGTACCCTCTGCACGCCATGGACG	*Nco*I	61°C	59°C
cas1/*2* rev	CGTATTAGGATCCGGTTAATGCAGAGCG	*Bam*HI	56°C	
casa1 for, 1	CATCAAAACCATATGTTGACACTCCTGG	*Nde*I	53°C	52°C
casa1 rev, 2	ACTTACATGTGGAGTGCGCCAGA	-	52°C	
casa1 for, 3	TCTGGCGCACTCCACATGTACGT	-	54°C	57°C
casa1 rev, 4	GGAGTTTAGGCCGGATCCGTGAGG	*Bam*HI	57°C	

The first expression experiments carried out under standard growth conditions in LB medium at 37°C resulted in a very low protein yield of Cas1/2 in the soluble and the

pellet fraction. After cell disruption and heat precipitation only a faint protein band at 29 kDa could be identified in the SDS-gel (s. 2.7.4). Also modifications such as different host strains or variation of cultivation temperatures did not result in higher recoveries (Fig. 3.19). However, a better production of protein was observed by changing the *E. coli* medium (NZA and 2x YT). Furthermore, the expression was improved by decreasing the temperature to 20°C after 0.1 mM IPTG induction and further incubation overnight.

Under these conditions the Cas1/2 protein was expressed, identified in the pellet fractions by its electrophoretic mobility (Fig. 3.19). One reason could be the amino acid composition of Cas1/2, since many positively charged amino acids are incorporated within the protein (47 of 264 amino acids: 18 % arginine and lysine). Probably, the supply of these amino acids was improved in the different media allowing a higher protein synthesis. Consequently, for following expressions of Cas1/2 the NZA medium was used. Expression studies of CasA1, also in NZA medium, resulted in a very low protein yield of the three proteins Cas4, Cas1/2 or Csa1 in the soluble and the pellet fraction (data not shown).

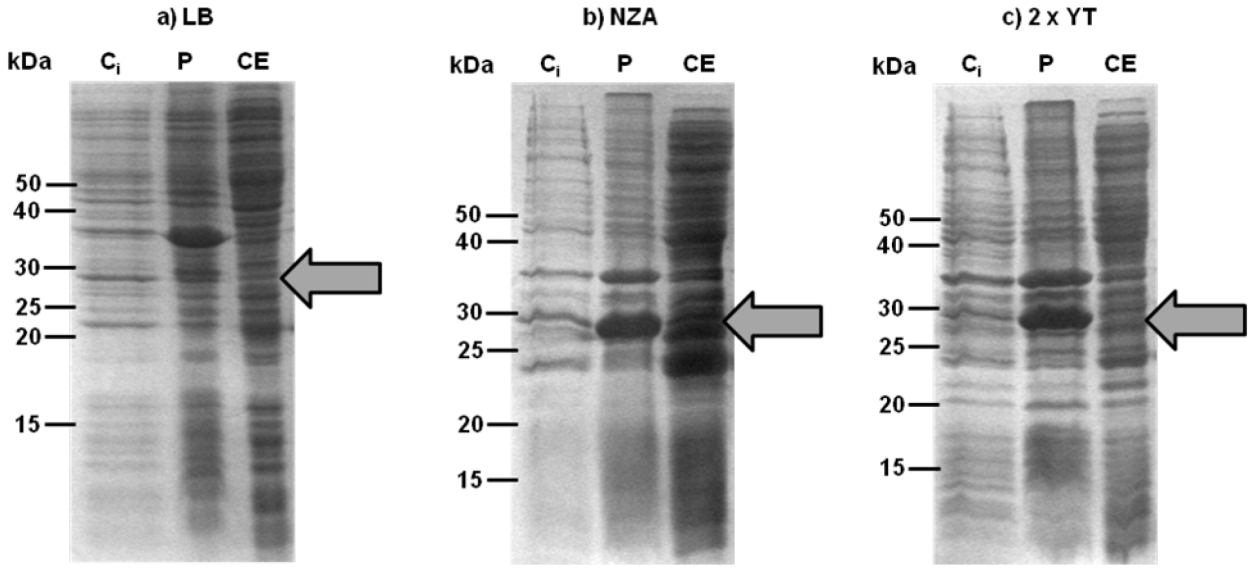


Fig. 3.19: Expression of Cas1/2: Influence of the culture medium. Expression of Cas1/2 (29 kDa) in *E. coli* Rosetta(DE3) in different media. **a)** LB medium (1 % peptone, 0.5 % yeast extract, 1 % NaCl, pH 7.2). **b)** NZA (1 % N-Z-Amine A®, 0.5 % yeast extract, 1 % NaCl, pH 7.2). **c)** 2 x YT (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl (w/v), pH 7.2). After reaching OD_{600} = 0.6, cultures were cooled down to 20°C and induced with 0.1 mM IPTG o/n. 1 g cells were disrupted by French press and homogenates separated on a 15 % SDS-gel. M: Unstained Protein Ladder, C_i: Lysed *E. coli* cells after 3 h IPTG induction, P: pellet fraction after centrifugation of cell extracts, CE: crude extract, soluble supernatant after centrifugation of cell extracts.

3.4.3 Refolding of CasA1 complex

To study the function of the proteins (Cas4, Cas1/2, Csa1) encoded by the operon *casa1*, their native structure had to be reconstituted, i.e. the aggregated state of these proteins had to be resolved and transformed into a soluble state. However, all trials to unfold the proteins separately by denaturing agents, such as 6 M GuHCl or 8 M urea, and refold them separately by diluting in refolding buffer, failed.

Therefore, an alternative method based on a synergistic reconstitution by mixing equal amounts of the three unfolded proteins in 4 M GuHCl was applied. This procedure allows proteins to reconstitute by rapid dilution in GuHCl-free refolding buffer (s. 2.7.2.5). Afterwards, the solution was centrifuged to separate aggregated (pellet) from refolded (i.e. soluble protein in the supernatant) proteins. Furthermore, the supernatant was concentrated to 1 ml (Vivaspin 20, size exclusion: 10,000 MW), analysed via SDS-PAGE and protein concentrations determined (s. 2.7.3 and 2.7.4). Although the presence of Mg^{2+} improved the refolding, the highest recovery of soluble protein was obtained in the presence of 500 mM arginine. As compared to parallel assays consisting only of Cas4 and Csa1 (partial complex, pCasA1), the recovery of soluble protein was by three- to fivefold higher in the presence of all three proteins (Fig. 3.20). Overall, 1 mg recombinant CasA1 complex was obtained from 1 g cells. From that it must be concluded that the three proteins encoded by the operon *casa1* strongly interact in the refolding process with highest efficiency in the presence of all three proteins.

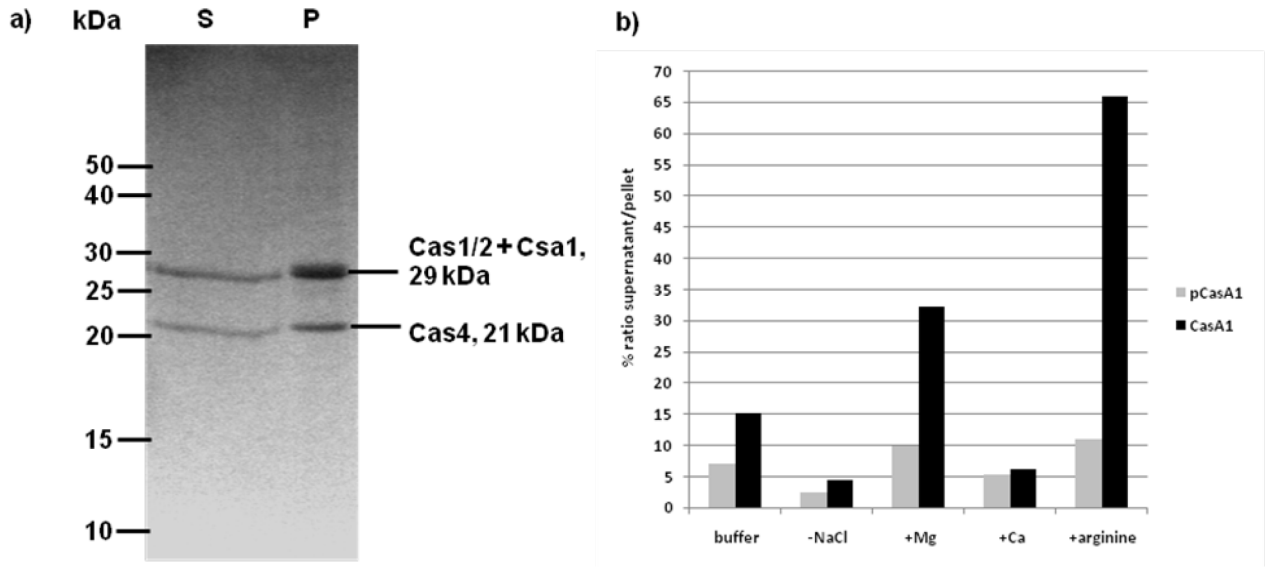


Fig. 3.20: *In vitro* reconstitution of the CasA1 complex. Unfolded proteins Cas4, Cas1/2 and Csa1 are combined and simultaneously (i.e. in one assay) refolded to CasA1 by rapid dilution. **a)** SDS gel electropherogram of the reconstitution assay (refolding buffer + 500 mM L-arginine), S: supernatant fraction (refolded complex), P: pellet fraction (aggregated protein), marker: Unstained Protein Ladder. **b)** ratio (%) between complete protein content of supernatant and pellet in different refolding buffers: buffer (40 mM Tris/HCl pH 7, 10 mM ß-Me, 10 % glycerol, 300 mM NaCl), -NaCl (buffer without NaCl), +Mg (buffer plus 10 mM $MgCl_2$), +Ca (buffer plus 10 mM $CaCl_2$), +arginine (buffer plus 500 mM L-arginine), CasA1: formation of the complete complex, pCasA1: formation of the partial complex.

3.4.4 Functional analyses of CasA1 complex

Since sequence comparisons show significant similarities of Cas4, Cas1/2 and Csa1 with known nucleases, nuclease activity tests were performed with specific CRISPR RNA as substrate. For long RNA transcripts the CRISPR locus TTX_4 was selected and primer constructed covering the whole sequence with a length of 2192 bp. The specific CRISPR RNA substrates were produced by PCR Taq-polymerase amplification with the primer sets *TTX_4* for and T7-rev, in which the rev primer contained the T7-polymerase binding site for the subsequent *in vitro* transcription (s. 2.5.6). The CasA1 complex was generated by refolding Cas4, Cas1/2 and Csa1 in the presence of 500 mM L-arginine as mentioned above (s. 2.7.2.5). Increasing protein concentrations were incubated with 40 ng CRISPR RNA for 30 min at 50°C in reaction buffer (s. 2.5.9). Afterwards, protein was removed from all samples by LiCl-precipitation and the reaction products were separated either on a 1 % formaldehyde-agarose gel or a 12 % denaturing PAGE (s. 2.5.4). As shown in Fig. 3.21, the complex CasA1 showed

RNase activity at protein concentrations of 12.5-17 ng/µl, visible by smaller RNA products.

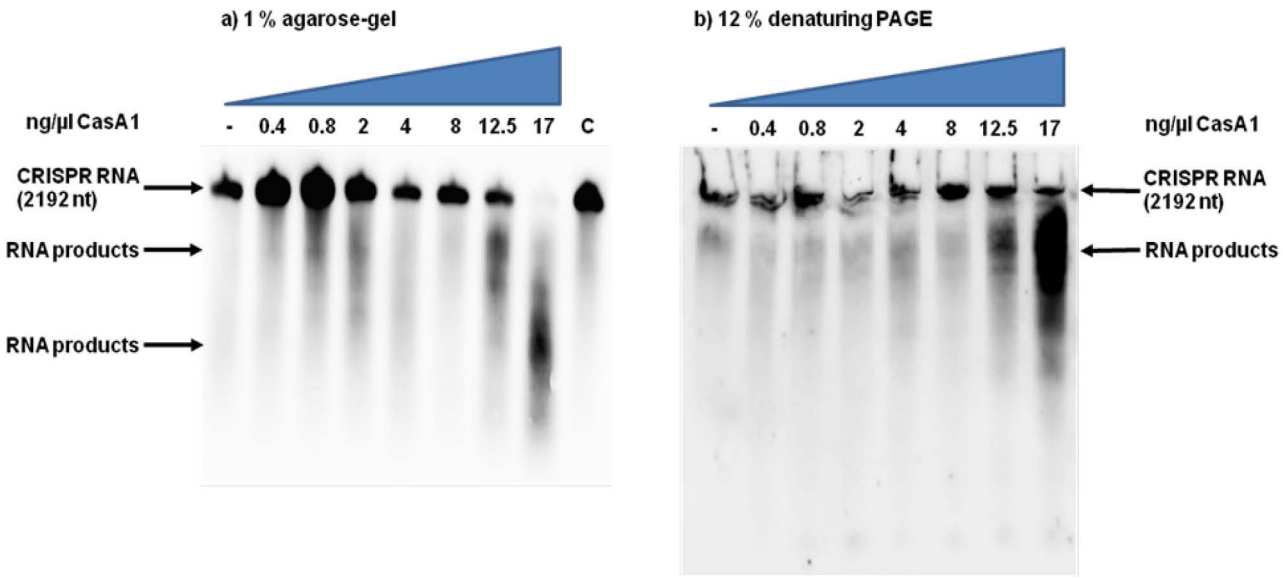


Fig. 3.21: Test for RNA nuclease activity of CasA1. 40 ng specific CRISPR RNA (TTX_4 cluster, 2192 nt) incubated at 50°C for 30 min with different concentrations (0-17 ng/µl protein) of CasA1, all samples were LiCl-precipitated after incubation. **a)** Reaction products separated on 1 % formaldehyde-agarose gel, C: refolding buffer without protein CasA1. **b)** Reaction products separated on 12 % denaturing PAGE.

3.5 Analyses of the archaeal CRISPR-associated complex 2 (CasA2)

3.5.1 Genomic context and sequence information of *casa2*

Similar to the described gene assembly of the *casa1* operon also the organisation of the genes *csa5*, *csa2*, *cas5a*, *cas3*, *cas3hd* and *csa4* (TTX_1250-1255) show characteristics of an operon: Close assembly of genes with partial overlaps of their gene sequences and characteristic promoter motifs (BRE-site and TATA-box) in front of the *csa5* gene, the first gene of the operon (*casa2*). The formation of a polycistronic mRNA, comprising the transcripts of all six genes, confirms the structural and functional unit of an operon. However, detailed functional analyses of the encoded proteins are still missing. Only sequence comparisons with known proteins allow some predictions of their function.

The sequence based function predictions for the first two encoded proteins of *casa2* revealed only basic structural motifs. Thus, the Csa5 protein (archaeal Cas protein) contains a coiled-coil motif (Coils) and Csa2 shows some homology to the DNA-binding regulator protein DevR from *Mycobacterium tuberculosis*, however without typical DNA-binding motifs. The Cas5a protein is related to the RAMP superfamily, a group of proteins with predicted RNA-binding activity (Makarova *et al.*, 2006). The Cas3 protein belongs to the family of DEAD/DEAH-box helicases, also found in other crenarchaeal organisms. The structure of these proteins is characterised by 2 domains, a helicase domain and a putative HD-type phosphohydrolase/nuclease domain, separated in two ORFs *cas3* and *cas3hd*. A noticeable feature of the *T. tenax* Cas3HD is a leucine-rich repeat (LRR, 2Zip) motif. LRRs are forming arc or horseshoe shapes and appear to provide for the formation of protein-protein interactions (Enkhbayar *et al.*, 2004). No functional predictions could be made for Csa4.

3.5.2 Cloning and heterologous expression of genes assembled in *casa2*

To analyse the function of the proteins encoded by *casa2*, the respective genes of *casa2* were singly cloned and heterologously expressed in *E. coli*. The genes were amplified by PCR mutagenesis with *Pfu* DNA-polymerase and the specific primer sets (primer: Tab. 3.13, PCR settings: Tab. 3.14). Additionally, the *cas3* gene was cloned in frame with a C-terminal His-tag (6x histidine).

Tab. 3.13: Primer sets for amplification of *casa2* genes. Primer sets used in *Pfu*-polymerase PCR mutagenesis assays for introduction of restriction sites and $(GTG)_6$ for adding His-tag in Cas3. Restriction sites (rs) underlined, codons for His-tag marked in black

Name	Sequence (5' - 3')	Rs	T_m
cac for	TGTATTCTCTTCCATATGGAGTCT	*Nde*I	47°C
cac rev	CGGCTCTAAGCTTGGGGGGAGCCA	*Hind*III	59°C
csa2 for	TGCTTAAGGGCATATGAGGGTGGC	*Nde*I	54°C
csa2 rev	GGCGCCCTGGCCTGGATCC	*Bam*HI	57°C
cas5a for	GGCAACCCGCTAGGAGCATATGCACTAC	*Nde*I	59°C
cas5a rev	CGCCAGGCTCGGATCCCTCACC	*Bam*HI	59°C
cas3 for	GGCTGTGTTCCCTGAGGCATATGGTT	*Nde*I	56°C
cas3 rev	CGGAAGTAGGCGCAGGAATTCATCG	*Eco*RI	56°C
cas3 his rev	GTAGGCGCAGGAATTCAGTGGTGGTGGTGGTGGTGTCGCCTCCTC	*Eco*RI	70°C
cas3hd for	AAGGGGGCGGAGGAGCATATGAG	*Nde*I	56°C
cas3hd rev	GCGTGATCATCTCCTGGATCCACTG	*Bam*HI	56°C
csa4 for	AGTGAGCGAGCTGGGCATATGAC	*Nde*I	52°C
csa4 rev	CGTGGGGCAGGATCCGAATTT	*Bam*HI	51°C

For expression using the pET system, the amplified genes were cloned into pET24a (s. 2.6.8.3). The recombinant vectors were subsequently used in transformation of *E. coli* Rosetta(DE3) (s. 2.7.1).

Tab. 3.14: PCR settings for amplification of *casa2* genes. PCR settings for *Pfu*-polymerase PCR mutagenesis assays (s. 2.6.7.3), elongation at 72°C.

Product	Primer sets	T_m PCR	Elongation	Length (bp)
cac	*cac* for / *cac* rev	48°C	1:00 min	393
csa2	*csa2* for / *csa2* rev	59°C	2:00 min	990
cas5a	*cas5a* for / *cas5a* rev	59°C	1:30 min	681
cas3	*cas3* for / *cas3* rev	63°C	4:00 min	1701
cas3 his	*cas3* for / *cas3* his rev	65°C	4:00 min	1719
cas3hd	*cas3hd* for / *cas3hd* rev	58°C	1:30 min	681
csa4	*csa4* for / *csa4* rev	57°C	2:30 min	1086

The purification of the expressed proteins revealed well-known problems, because most of the six proteins (Cas5a, Cas3, Cas3HD and Csa4) were expressed as inclusion bodies. Not enough protein was obtained in soluble fractions and that was also the case for the His-tagged Cas3 protein with Ni-NTA agarose (s. 2.7.2.7). As an example, the SDS-gel electropherogram of Cas3 is depicted showing the respective protein (Cas3: 63 kDa) in the pellet fraction after centrifugation of crude extracts (Fig. 3.22).

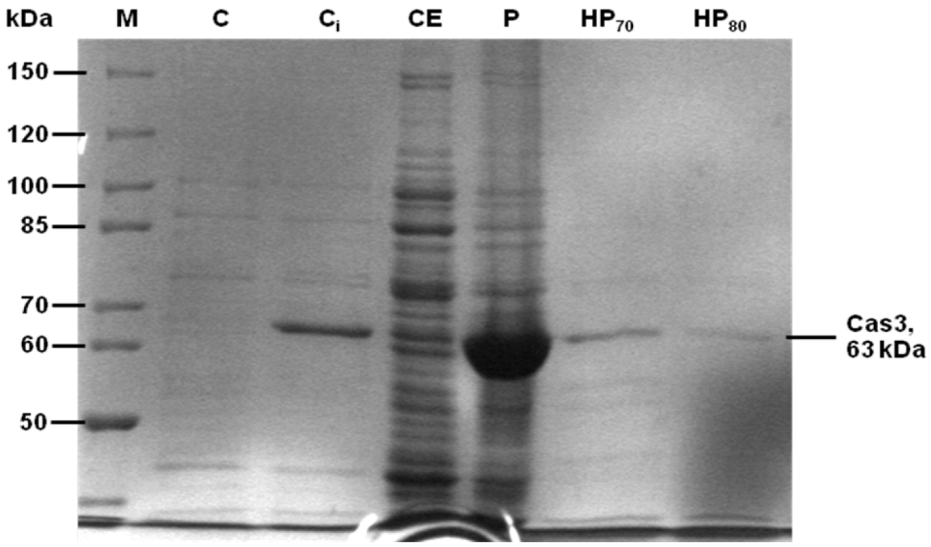


Fig. 3.22: Purification of the recombinant Cas3 protein. Electropherogram of a 7.5 % SDS-PAGE, M: Unstained Protein Ladder, C: Lysed *E. coli* cells before IPTG induction, C_i: Lysed *E. coli* cells after 3 h IPTG induction, CE: crude extract, soluble fraction after centrifugation of cell extracts, P: pellet fraction after centrifugation of cell extracts, HP_{70-80}: heat precipitated and centrifuged crude extract at 70 or 80°C.

As an exception, the expressed proteins (Csa5 and Csa2) of *casa2* were soluble and could be directly purified. Both proteins were extracted by cell disruption and enriched in the heat precipitation step at 80°C. Csa5 was further enriched by anion exchange chromatography on Q-sepharose, Csa2 by cation exchange chromatography with heparin-sepharose (s. 2.7.2.1 and 2.7.2.2). In a SDS-gel electropherogram the corresponding fractions containing Csa5 (15 kDa) or Csa2 (36 kDa) were identified, pooled and concentrated (s. 2.7.4). 5 mg recombinant Csa5 and 1.5 mg Csa2 were obtained from 1 g cells. Fig. 3.23 documents the different purification steps for both proteins.

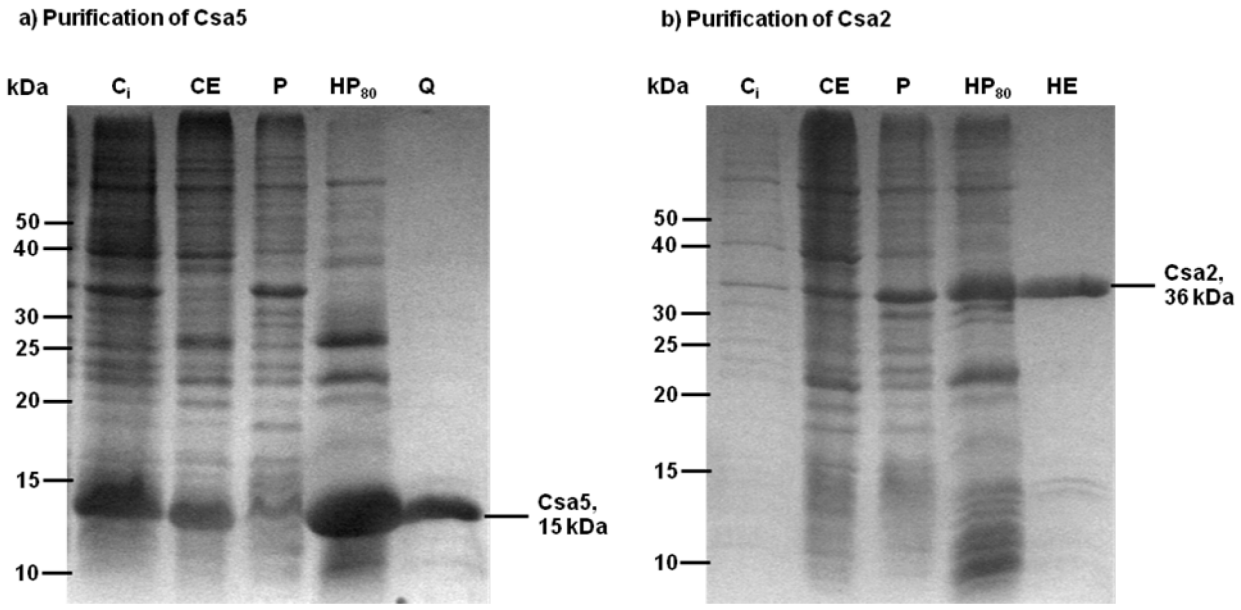


Fig. 3.23: SDS gel electropherogram documenting the purification of the recombinant Csa5 and Csa2 proteins from *T. tenax*. a) Purification of Csa5: 15 % SDS-PAGE, M: Unstained Protein Ladder, C_i: lysed *E. coli* cells 3 h after IPTG induction, CE: crude extract, soluble supernatant after centrifugation of cell extracts, P: pellet fraction after centrifugation of cell extracts, HP_{80}: heat precipitated and centrifuged crude extract at 80°C, Q: pooled and concentrated fractions after Q-sepharose. **b)** Purification of Csa2: same description as a), except for HE: pooled and concentrated fractions after heparin-sepharose.

3.5.3 Refolding of CasA2 complex

Similar to the reconstitution of CasA1, the four proteins of CasA2 (Cas5a, Cas3, Cas3HD and Csa4) were expressed as inclusion bodies in *E. coli*, unfolded separately and tried to refold them in the presence of their partners expecting that also these proteins refold synergistically forming a protein complex comparable to CasA1 consisting however of six Cas protein (termed “CasA2”).

First of all, the aggregated proteins Cas5a, Cas3, Cas3HD and Csa4 were extracted by disruption of 5 g recombinant *E. coli* cells and their inclusion bodies were purified (s. 2.7.2.4). Next, the recombinant proteins were solubilised in 6 M GuHCl (protein concentrations 0.6-1.3 mg/ml). The *in vitro* reconstitution of the CasA2 complex was carried out by a stepwise dilution of the denaturing agent GuHCl (s. 2.7.2.6). Equal amounts (100 µg) of the solubilised proteins and the purified proteins were refolded together by mixing the protein solutions in a GuHCl-containing buffer, filled them in a dialysis bag and decreased stepwise the concentration of GuHCl in the surrounded buffer. Afterwards, the solution was centrifuged to separate aggregated (unfolded

protein in the pellet) and soluble protein (refolded in the supernatant). The supernatant was concentrated to 800 µl (Vivaspin 20, size exclusion: 10,000 MW), analysed via SDS-PAGE and its protein concentration determined (s. 2.7.3).

Refolding of one (p1CasA2: refolding of Cas3 or Cas3HD) or four proteins (p4CasA2: refolding of Cas5a, Cas3, Cas3HD and Csa4) resulted in very low amounts of refolded protein in the supernatant (5-10 %). In comparison to the full CasA2 complex comprising all six proteins, the partial complex was refolded with a two-fold decreased efficiency (Fig. 3.24). To optimise the reaction conditions, buffer composition and addition of low and high molecular weight compounds was varied and the reconstitution efficiency for the complete complex (CasA2) determined. The refolding experiments revealed that the addition of up to 6 µg/ml *T. tenax* total RNA (total RNA prepared by isopropanol precipitation, s. 2.5.2.1) enhanced the yield of the soluble fraction by factor two. The combination of total RNA and $CaCl_2$ in the refolding assay showed the best results with a yield of nearly 50 % soluble protein and a 1:1 ratio of the six proteins (Fig. 3.24).

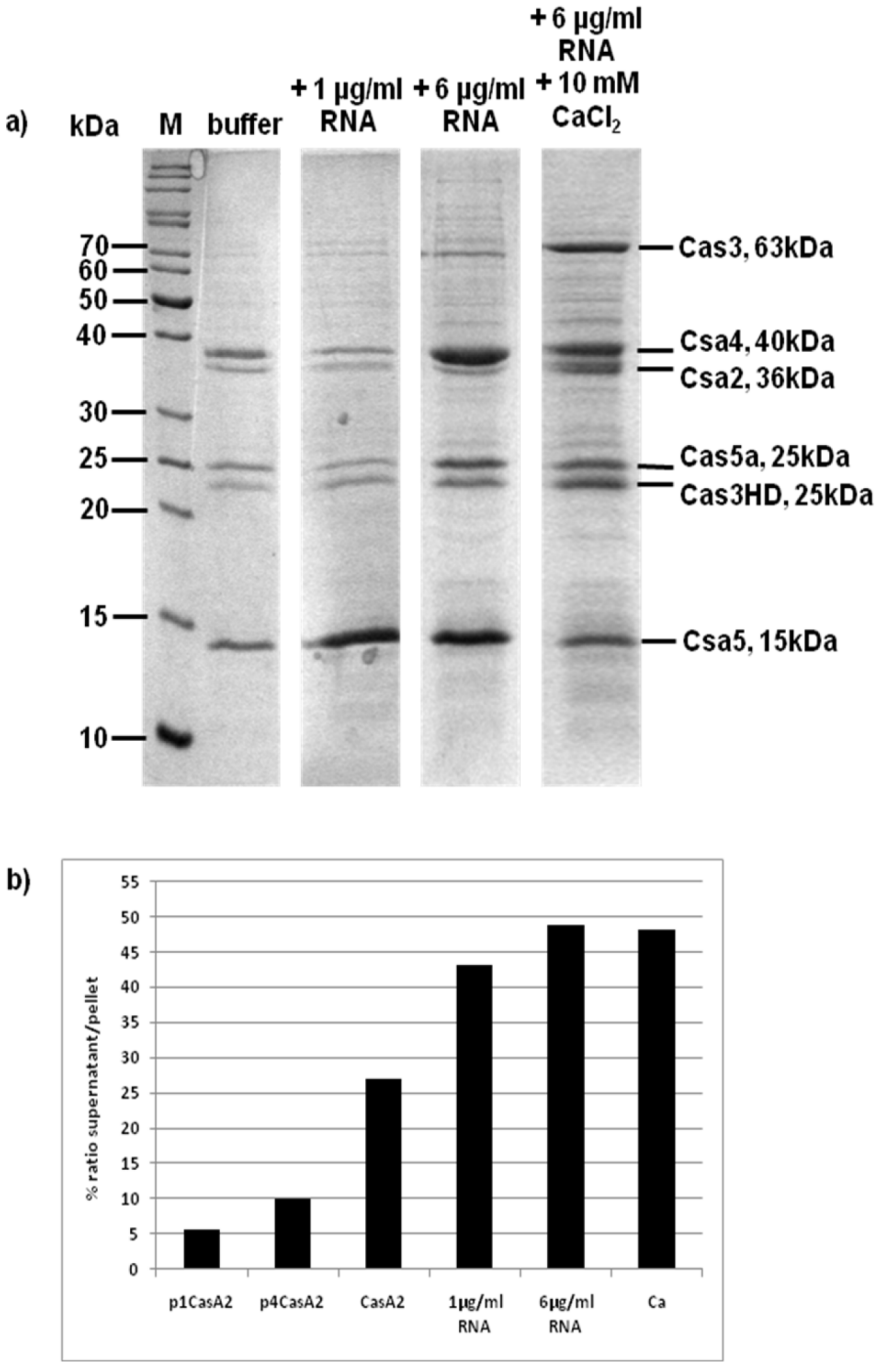


Fig. 3.24: *In vitro* reconstitution of the CasA2 complex. The separately unfolded proteins Cas5a, Cas3, Cas3HD and Csa4 and soluble purified proteins Csa5 and Csa2 were reconstituted to CasA2 by stepwise dialysis to remove GuHCl. **a)** SDS-gel electropherogram for identification of the constituents of various reconstitution assays for CasA2: in refolding buffer, + 1 µg/ml total RNA, + 6 µg/ml total RNA, + 6 µg/ml total RNA and 10 mM $CaCl_2$; shown are the supernatant fractions of the various assays containing the refolded proteins, marker: Unstained Protein Ladder. **b)** % ratio (protein content) of supernatant to pellet in different refolding buffers (native buffer: 100 mM HEPES/KOH pH 7, 10 % glycerol, 300 mM NaCl, 10 mM ß-Me; for refolding of p1CasA2, p4CasA2 and CasA2). Added 1 µg/ml total RNA, 6 µg/ml total RNA or 6 µg/ml total RNA and 10 mM $CaCl_2$ (*T. tenax* RNA prepared by isopropanol precipitation). Partial complexes: p1CasA2 (one protein: Cas3 or Cas3HD) and p4CasA2 (four proteins: Cas5a, Cas3, Cas3HD, Csa4).

Furthermore, the refolded CasA2 complex was analysed on a gel filtration column in order to determine its molecular mass and stability (s. 2.7.5). This experiment was performed with CasA2 complexes reconstituted with all six proteins in presence of 6 µg/ml total *T. tenax* RNA. 200 µl protein solution (~75 µg protein content) was applied on the gel filtration column. Afterwards the collected fractions (volume: 450 µl) were concentrated by acetone precipitation and its composition was identified by SDS-PAGE (15 % PAA). The results indicated that the CasA2 complex exhibited considerable stability in the course of gel filtration (Fig. 3.25). In comparison to reference proteins with known molecular weights the largest fractions of CasA2 eluted between Ferritin (MW 443,000) and ADH (MW 148,000) representing a approximate molecular mass of 300 kDa. But the refolding process generated also some intermediates, representing smaller protein complexes composed of less than six CasA2 members or without RNA (data not shown).

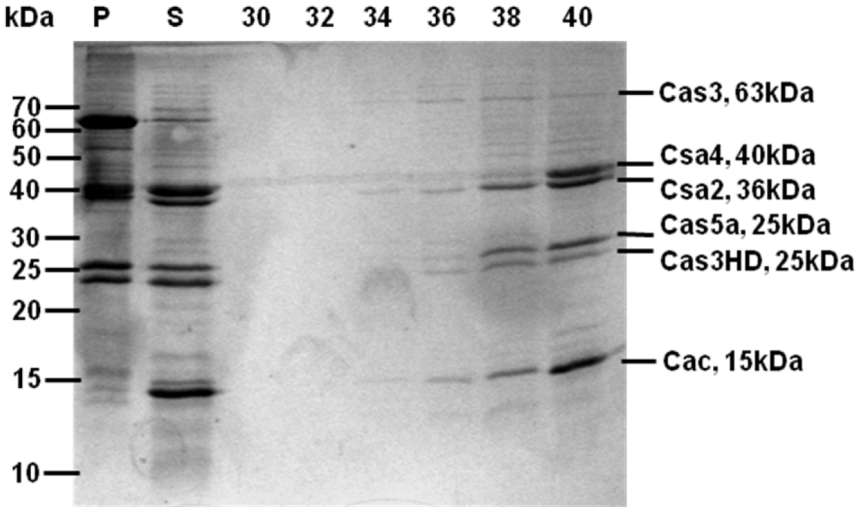


Fig. 3.25: SDS gel electropherogram of the refolded CasA2 complex separated by gel filtration on superose™. 15 % SDS-PAGE, M: Unstained Protein Ladder, P: pellet fraction, S: concentrated supernatant, 30-40: acetone precipitated fractions separated on gel filtration. 75 µg CasA2 were applied to the column, flow rate: 0.3 ml/min, column: Superose™ 6 10/300 (volume: 24 ml, length: 30 cm, diameter: 0.5 cm, Pharmacia).

3.5.4 Functional analyses of CasA2 complex

For studying the molecular function of the CasA2 complex, investigations with nuclease assays, motivated by sequence comparisons which suggest a nuclease activity of Cas3HD. The protein complex was not only incubated with specific CRISPR RNA of locus TTX_4 (2192 nt), but also with a short substrate composed of one repeat

and one spacer sequence (spacer 4.22, length: 71 nt). The specific CRISPR RNA substrates were produced by PCR *Taq*-polymerase amplification with the respective primer set, in which the reverse primer contained the T7-polymerase binding site for the subsequent *in vitro* transcription (s. 2.5.6). The CasA2 complex was reconstituted with the addition of total *T. tenax* RNA (6 µg/ml) as mentioned above. Different protein concentrations were incubated with 40 ng short and long CRISPR RNA for 30 min at 50°C in the reaction buffer (s. 2.5.9). Afterwards, protein was removed from all samples by LiCl-precipitation and the reaction products were analysed either by electrophoresis on a 1.5 % formaldehyde-agarose gel (for separation of long RNA; s. 2.5.4.1) or on a 15 % denaturing PAGE (for separation of short RNA; s. 2.5.4.2). The gels were blotted overnight and the reaction products detected (s. 2.5.8). As shown in Fig. 3.26, neither long nor short ssRNA was degraded by the CasA2 complex at concentrations up to 50 ng/µl protein. Quite obviously, for the complex CasA2 an ssRNA nuclease activity could be excluded under the chosen conditions.

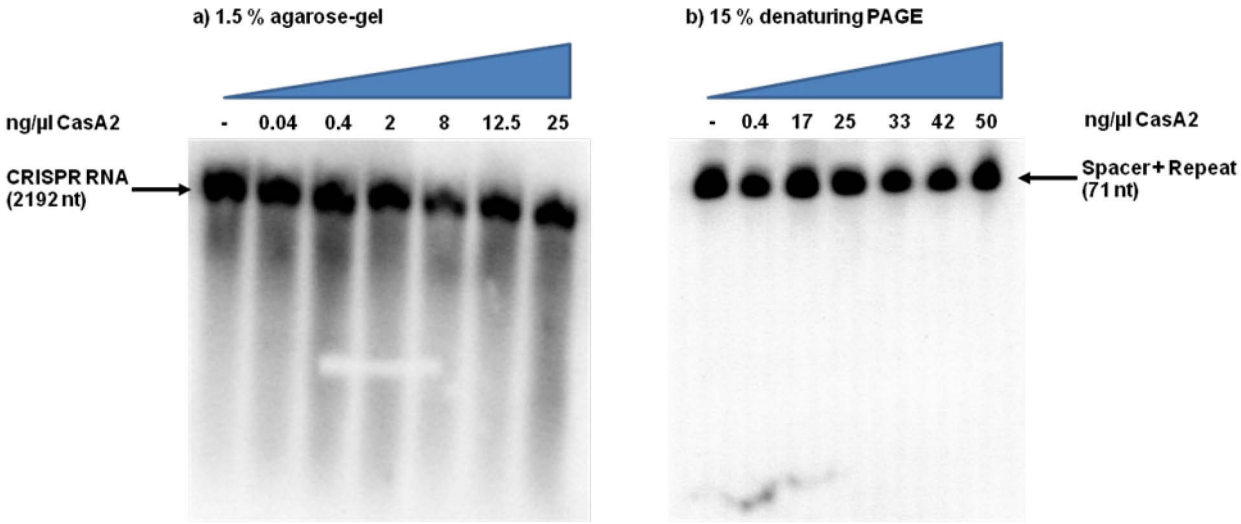


Fig. 3.26: Assay for RNA nuclease activity of CasA2 with long and short RNA. Incubation of 0-50 ng/µl CasA2 for 30 min at 50°C with 40 ng CRISPR RNA (long: TTX_4 cluster, 2192 nt or short: repeat + spacer 4.22, 71 nt); all samples LiCl precipitated after reaction. **a)** Reaction products of long RNA separated on 1.5 % formaldehyde-agarose gel. **b)** Reaction products of short RNA separated on 15 % denaturing PAGE.

Motivated by refolding experiments for the reconstitution of CasA2 indicating an interaction of CasA2 proteins with RNA, specific RNA binding studies with CasA2 were performed. As a target the short DIG-labelled CRISPR RNA (repeat + spacer 4.22) was used. 50 ng of the DIG-labelled probe were incubated with 0.05 to 15 ng/µl refolded CasA2 complex at 37°C or 30 min in binding buffer (s. 2.5.10). The RNA-protein complexes were separated from free RNA by a 6 % non-denaturing PAGE followed by blotting and detection by chemiluminescence (s. 2.6.5.2). Analyses by EMSA clearly indicated that the protein complex binds short CRISPR RNA substrate at least at protein concentrations of 0.5 ng/µl and above (Fig. 3.27).

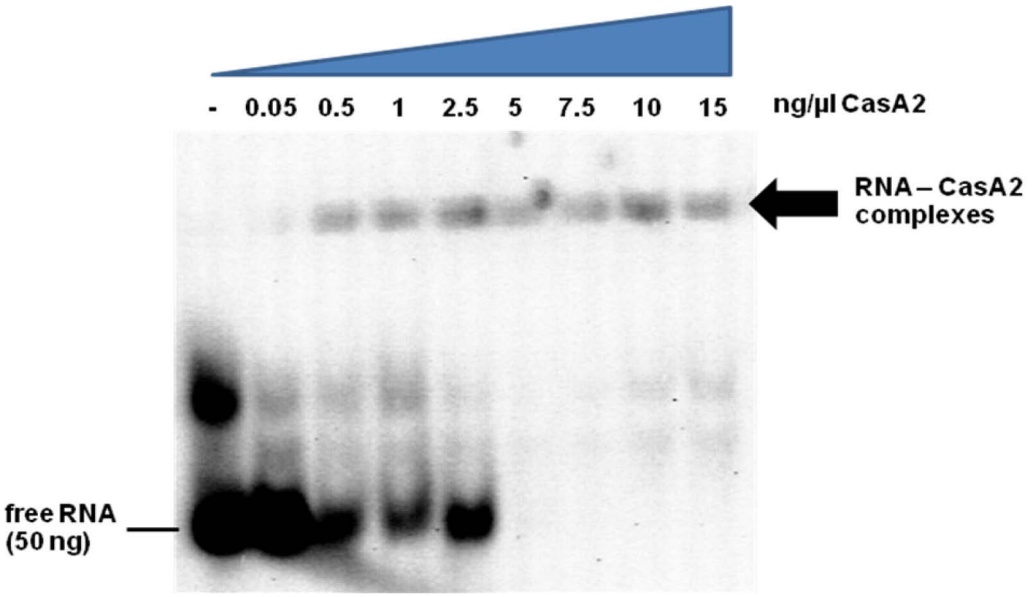


Fig. 3.27: EMSA with CasA2 and short CRISPR RNA (repeat + spacer 4.22). Different concentrations of CasA2 (0-15 ng/µl) incubated with 50 ng DIG-labelled RNA probe at 37°C. Putative RNA-protein complexes are marked by an arrow.

4. DISCUSSION

4.1 CRISPR loci

Clustered regularly interspaced short palindromic repeats (CRISPR) found in prokaryotes are non-contiguous direct repeats with a length of 24-48 nt. The sequence repeats are weakly palindromic at the 5'- and 3'-termini and separated by variable spacer sequences of similar size (Jansen *et al.*, 2002). CRISPR loci are flanked on one side by an AT-rich leader sequence of 200-350 bp length. In the course of this work seven CRISPR loci could be identifed in the genome of the hyperthermophilic Crenarchaeote *Thermoproteus tenax*. But the impact of CRISPR and the adjacent CRISPR-associated genes (*cas* genes) on prokaryotes is so far only poorly understood. The present work should contribute to a better understanding of molecular functioning of the CRISPR/Cas system in prokaryotic organisms.

4.1.1 Features of CRISPR in *Archaea*

The existence of CRISPR sequences is widespread in the two prokaryotic domains. CRISPR loci were identifed in most archaeal genomes, in 40 % of the bacterial genomes and on some plasmids (Grissa *et al.*, 2007). CRISPR have been firstly discovered in the *E. coli* genome and initially considered as some kind of noncoding repetitive elements (Ishino *et al.*, 1987). Due to the growing number of sequenced genomes, CRISPR were also found in other *Bacteria* and *Archaea* (Mojica *et al.*, 1995; Masepohl *et al.*, 1996; Jansen *et al.*, 2002; Fadiel *et al.*, 2003), but not in viruses or eukaryotic genome sequences. Strikingly, *Archaea* possess more and larger CRISPR compared to *Bacteria*. In extremely thermophilic and hyperthermophilic *Archaea* up to 1 % of the genome is covered by CRISPR (Nelson *et al.*, 1999; Lillestol *et al.*, 2006).

The seven identified CRISPR loci of *T. tenax* have different sizes, ranging from 7 to 34 repeats and comprehend 143 unique spacer sequences. Because of different similarities of repeat sequences, the seven CRISPR loci were separated into two groups (I and II). Interestingly, 75 % of the *T. tenax* spacers had a length of 41 to 46 bp. So, it seems to be a preferential feature for organisms to have a distinct length or structure of spacer sequences (Fig. 3.1). The comparison of 24 archaeal genomes shows typical features of archaeal CRISPR (Tab. 4.1). In general, all analysed genomes contain more than one repeat cluster. The highest number (20 CRISPR) could

be found up to now in *M. jannaschii*. The direct repeats are similar within a single locus, but different between CRISPR loci of the same genome. Thus, the repeats can be classified in groups dependening on disparities in sequences. Interestingly, cren- and euryarchaeal organisms differ significantly in sizes (24-25 vs. 28-37) and sequences of repeats (Plagens *et al.*, 2010a; in preparation).

Tab. 4.1: Comparison of 24 archaeal CRISPR arrays. Listed are the number of clusters, different repeat groups within the genome, length of repeats, number of spacers and the optimal growth temperature. CRISPR data are obtained from CRISPRdb, and optimal growth temperatures from the DSMZ.

Organism	CRISPR no.	Repeat groups	Length (nt)	Spacers no.	T_{opt}°C
Crenarchaeota					
Aeropyrum pernix	3	2	24-25	85	90
Staphylothermus marinus	11	2	25	108	90
Metallosphera sedula	4	2	24-30	373	65
Sulfolobus solfataricus	7	1	24-25	416	85
Sulfolobus tokodaii	5	2	24-25	457	75
Sulfolobus islandicus L.S.2.15	4	2	24-25	232	80
Thermofilum pendens	10	3	24-25	147	88
Pyrobaculum aerophilum	6	2	25	130	98
Pyrobaculum arsenaticum	4	2	24	127	95
Pyrobaculum caildifontis	7	2	24	90	90
Thermoproteus neutrophilus	10	5	24-25	224	85
Thermoproteus tenax	7	2	24-25	143	86
Euryarchaeota					
Archaeoglobus fulgidus	3	2	30-37	149	85
Haloarcula marismortui pNG400	3	2	30	80	37
Methanosphaera stadtmanae	3	2	29-30	121	37
Methanocaldococcus jannaschii	20	3	29-30	178	85
Methanospirrilum hungatei	7	3	27-37	264	37
Methanococcoides burtonii	2	2	29-30	84	23
Methanosarcina acetivorans	5	3	30-37	69	40
Methanosarcina barkeri	4	4	30-39	95	37
Methanosarcina mazei	4	2	24-35	128	37
Pyrococcus furiosus	7	2	28-30	200	100
Thermococcus kodakaraensis	3	1	30	74	85
Picrophilus torridus	4	2	28-30	116	55

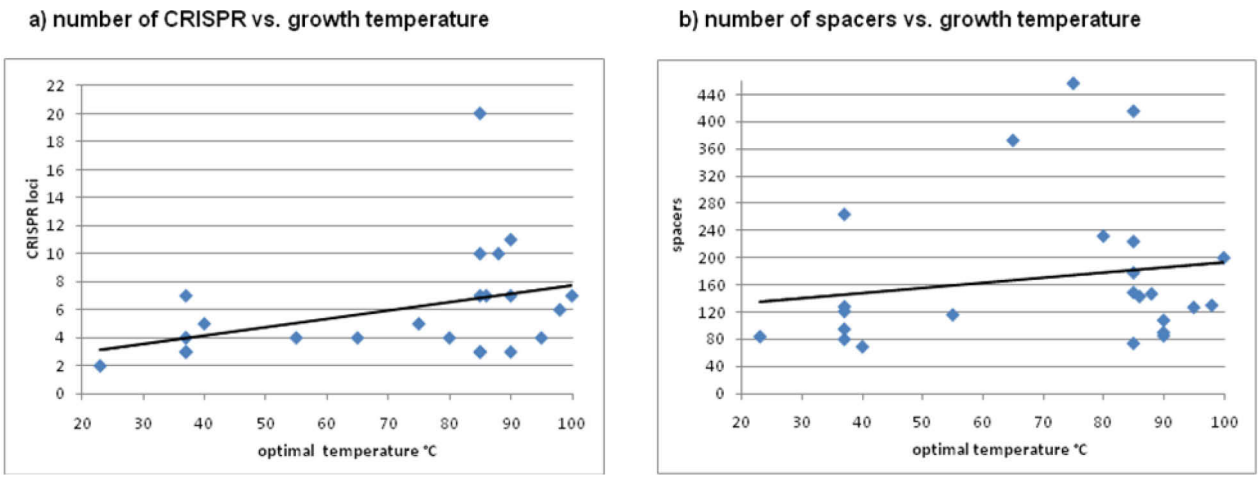


Fig. 4.1: Correlation of number of CRISPR or spacers and optimal growth temperatures in 24 archaeal organisms. Genomes listed in Tab. 4.1 analysed for number of CRISPR and spacers, identified by CRISPRdb and optimal growth temperatures (source: DSMZ).

Interestingly, some correlation of number of CRISPR arrays and number of spacers with the optimal growth temperatures of archaeal organisms was observed (Fig. 4.1). Thus, it seems that *Archaea* with a higher optimal growth temperature comprehend generally more CRISPR and spacers.

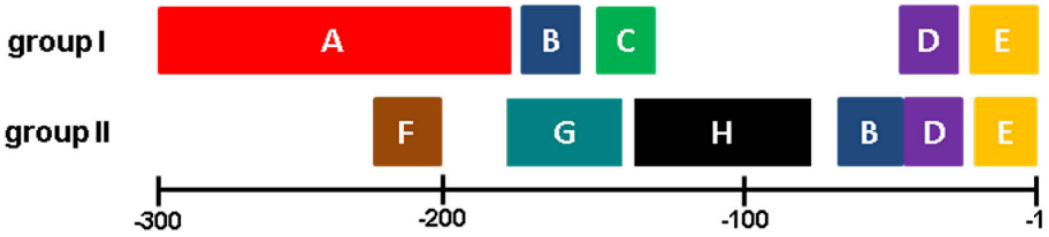
4.1.2 Conservation of leader sequences in *T. tenax*

Leader sequences, located directly upstream of the first repeat, share approx. 80 % sequence identity within a species, but show no obvious similarity with leaders of taxonomically unrelated species. Furthermore, leader sequences do not contain open reading frames and are only found upstream of a CRISPR locus and not elsewhere in the genome (Jansen *et al.*, 2002). So, CRISPR and leader sequences represent an inseparable unit.

In order to identify the leader sequences in *T. tenax*, both flanking sites of a CRISPR locus were compared with respect to their AT-content. The results show that the AT-content of one flanking site is at an average of 5 % higher than of the other flanking site (s. 3.1.1). Alignments of leader sequences in *S. solfataricus* revealed a high conservation and different motifs with a length of 30-100 bp. In different *Sulfolobus* species, the transcription start site of the CRISPR locus is located directly upstream of the first repeat, preceded by typical archaeal BRE-sites and TATA-box motifs (Lillestol *et al.*, 2009). Thus, the identified leader sequences in *T. tenax* were analysed by multiple sequence alignments (Clustal W; Larkin *et al.*, 2007) and the similarity of leader sequences was in accordance with the CRISPR groups I and II.

Both groups showed typical BRE-site and TATA-box motifs, but different motif blocks in all seven leader sequences (Fig. 4.2). Taken together, the identified leader sequences were confirmed by the high similarity within the two groups in *T. tenax* and the presence of typical archaeal-like transcription start sites.

a) Motif maps for leader regions



b) BRE/TATA motifs in all leader sequences

```
                                                    -30        -20        -10
TTX_6          ATCAACGCAAAACCCACCAGAGAAAAACTTAAAAACAGCCAAAAACCAAAACCCAGAAGGCCCG
TTX_7          ACTAACACAAAACCCACCAGAGAAAAACTTAAAAACAGCCAAAAGCCAAAACCCAGAAGGCCCG
TTX_4          ACTAACACAAAGACCCACAGAGGAAAACTTAAAAATCGGCAAAAACTAAAGACCAGAAGGCCCG
TTX_5          ACCAAGACAAAGACCCGCAGAGGAAAACTTAAAAATCGGCAGAAACTAAAGACCAGGAGGCCCG
TTX_1          ATCAACGTCAGTCTCGTGAGAAAAAAGCTTAAAAGCTCGTAGAAAC-AAAGACAACAATACCCG
TTX_2             GCCGGGCGGGCGGAGGATCTGGGCAAAATTTAAATAGCCGGGGCCGCAATCGACGTGGGCG
TTX_3             GCCGGGCGGGCGGAGGATCTGGGCAAAATTTAAATACCCGAGGCCGCAGTCAACGTGGGCG
```

Fig. 4.2: Multiple sequence alignments of leader sequences. a) Motif maps for leader sequences of group I and II CRISPR in *T. tenax*. A, G, H: complex similarity, B: 90 % GC-rich motif, C: region similar to repeat sequence, D: BRE-site and TATA-box, E: transcription start region, F: AT-rich motif. **b)** BRE- and TATA-motifs in every CRISPR, BRE-site underlined, TATA-box marked in purple.

4.1.3 Origin of spacers

Some spacer sequences show similarity to known viruses, plasmids, transposons or chromosomal elements (Bolotin *et al.*, 2005; Makarova *et al.*, 2006; Mojica *et al.*, 2009). On the basis of these observations, the scheme of a defense system against mobile genetic elements was established, as spacer sequences interfere with foreign sequences by complementarity (van der Oost *et al.*, 2009). Therefore, a CRISPR array would represent a "genetic memory" of previous attacks. So, the confirmation of spacer homologies is an important and revealing step towards the understanding of CRISPR functions. In many studies the authors tried to define the origin of spacers, but only for a limited number of sequences a firm identification could be performed. This is probably due to the fact that a vast amount of uncharacterized genetic elements is existing in nature.

In *Yersinia pestis* 24 spacers were found that are homologous to a defective lambdoid prophage (YPO) within the *Y. pestis* genome and seven matches were detected to other genes or intergenic regions (Pourcel *et al.*, 2005). From 15 analysed archaeal genomes only four organisms had significant matches of spacer sequences to foreign genetic elements (Mojica *et al.*, 2005), such as *M. thermautotrophicus* (9 of 169 spacers to phages) or *P. aerophilum* (1 of 129 spacers to chromosomal sequences). Eight different strains of *S. islandicus* contained 1393 spacers and hereof 233 significant BLAST matches were observed. 138 spacers correlated to viral genomes and 14 to plasmids, but only ten spacers were identified to be 100 % identical (Held and Whitacker, 2009; Shah *et al.*, 2009).

The analyses of spacers in *T. tenax* showed that one spacer showed complete identity to the *T. tenax* gene TTX_0660, which encodes for an adhesin-like protein with characteristic motifs for transmembrane-helices (Tab. 3.2). Further spacers matched the sequence of a cell division control protein Cdc21 and a creatinine amidohydrolase of *T. tenax*, with identities of approx. 50 %. Hits with similarities of approx. 40-50 % were detected for genes of several archaeal viruses, i.e. *Acidianus two-tailed virus* (ATV), *Acidianus filamentous virus 6* (AFV6) or *Pyrobaculum spherical virus* (PSV), but only PSV is known to infect *T. tenax* (Häring *et al.*, 2004). On the basis of the above mentioned hypothesis about the origin of CRISPR spacers, these observations could suggest that indeed these viruses or their relatives left their traces in *T. tenax* CRISPR loci upon previous infections. Other phages infecting *T. tenax*, e.g. *Thermoproteus tenax virus 1* (TTV1, Neumann and Zillig, 1990; Zillig *et al.*, 1996) or *Thermoproteus tenax spherical virus 1* (TTSV1, Ahn *et al.*, 2006) showed no homology to spacers of *T. tenax*. However, the origin of the vast majority of spacers is still unknown, possibly because relevant infective phages or their close relatives and other mobile genetic elements are not yet sequenced.

4.1.4 Transcription of CRISPR loci

Small noncoding RNAs (tRNA, rRNA, miRNA or siRNA) are present in all domains of life. They function in wide array of essential cellular processes, such as translation, mRNA splicing or gene regulation. In eukaryotes siRNAs and miRNAs act in post-transcriptional gene silencing by targeting exogenous or endogenous RNAs, known as the RNA-interference process (Hannon, 2002). Other classes of small RNAs in eukaryotes are piwi-interacting (piRNAs) or repeat-associated small interfering RNAs

(rasiRNAs), regulating the spreading of selfish genetic elements such as transposons or repeat elements (Aravin *et al.*, 2007; Hartig *et al.*, 2007). In prokaryotes many small RNAs were identified, in most cases trans-acting (intermolecular, e.g. 6S rRNA) or cis-acting (intramolecular, e.g. riboswitches; Waters and Storz, 2009) RNA molecules.

Strong evidence was provided that also CRISPR loci are transcribed and processed into a series of smaller RNAs, corresponding in length to multiple repeat-spacer units (Tang *et al.*, 2002, 2005). The transcription start site is located immediately upstream of the first repeat, preceded by archaeal BRE/TATA motifs within the leader sequence (Fig. 4.2 and Lillestol *et al.*, 2009). In *A. fulgidus* Northern blot analyses with probes covering the CRISPR repeat sequence (30 nt) revealed sizes of 272, 204, 136 and 68 nt (Tang *et al.*, 2002). Similar results were obtained in *S. solfataricus* with a 24 nt long repeat probe (540, 360, 180 and 60 nt; Tang *et al.*, 2005). These results suggest a biogenesis pathway for transcript products, as long primary CRISPR transcripts are endonucleolytically cleaved within repeats to produce small RNA molecules flanked by parts of repeat sequences at 5'- and/or 3'-end (Brouns *et al.*, 2008). Lillestol and co-workers (2009) identified for *S. acidocaldarius* by Northern blots with spacer probes also smaller products of 45 nt length indicating progressive exoribonuclease trimming. Additionally, they could detect transcripts not only of the sense strand, but also of the antisense strand. However, the anticipated double-stranded small CRISPR transcripts were not found. In *P. furiosus* 17.3 % of total small RNA (smaller 50 nt) consists of RNA from CRISPR loci and two-thirds derived from the first third of a cluster (Hale *et al.*, 2008). The transcripts covered the spacer sequence and variable amounts of the flanking repeat sequence (5'-end: 0-8 nt, 3'-end: 0-22 nt).

On the basis of these experiments, the results obtained for *T. tenax* can be further interpreted (Fig. 3.4). The smallest detectable transcripts of ~50 nt length were clearly larger then the basic spacer sequences of 43 bp, which showed that the RNA transcripts had to be flanked by parts of the repeat sequence (24 nt). Overall, the CRISPR RNAs showed also larger sizes of ~130 nt, ~110 nt and ~70 nt, assuming that large CRISPR RNA transcripts were stepwise endo- and exonucleolytically processed (Fig. 4.3). But for a clear definition of CRISPR transcript length and structure further experiments are needed. One approach could be cloning of total

RNAs smaller than 200 nt by reverse transcription and standard microRNA cloning protocols (Lau *et al.*, 2001). Noteworthy, no differences were detected between normal grown *T. tenax* cells and cultures treated with higher temperatures or UV-light. This indicates that the formation of small CRISPR RNAs is not regulated, but the product of constitutive transcription in *T. tenax*.

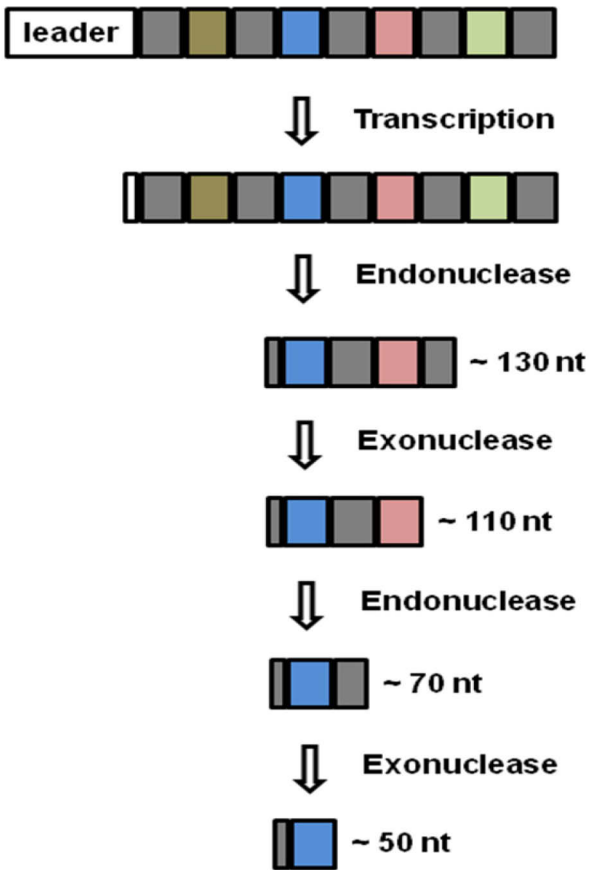


Fig. 4.3: Hypothesis for generating small CRISPR RNAs in *T. tenax*. On the basis of transcription results from other *Archaea* and detected small RNAs of *T. tenax* (~130, ~110, ~70 and ~50 nt), it is likely that CRISPR loci are transcribed, starting from the leader site and large CRISPR transcripts are further processed by endo- and exonucleases. Repeat sequences (24 nt) grey boxes, spacers (43 nt) different coloured.

In *S. acidocaldarius* the processing sites of CRISPR transcripts could be identified within the repeat sequence (Lillestol *et al.*, 2009). 12 archaeal organisms were selected and their CRISPR repeats compared to *S. acidocaldarius* to identify conserved motifs for the putative processing sites. As shown in Fig. 4.4, especially the 3‘-end of crenarchaeal repeats is nearly identical in all analysed organisms. The putative restriction sites in comparison to *S. acidocaldarius* are located 8 nt downstream of the

5œend (after the conserved C residue) and 8 nt upstream of the 3œend (after the conserved rT residues), suggesting that in all Crenarchaeota homologous nucleases detect the conserved pattern of the repeat sequence. It can be concluded that the repeat sequence of a CRISPR transcript is a potential target for nuclease degradation.

```
MSE_4          -GTTAATCTTCTATAGAGTTGAAAG   24
SIS_3          -GTTAATCTACTATAGAATTGAAAG   24
APE_3          GCATATCCCTAAAGGGAATAGAAAG   25
STO_8          -GATAATCCTTAATGGAATTGAAAG   24
SAC_4          GATGAATCCCAAAAGGGATTGAAAG   25
SSO_3          GATTAATCCCAAAAGGAATTGAAAG   25
PCA_6          -GAATCTCAAGAAGAGGATTGAAAG   24
TNE_10         -GAATCTCAAGTTGAGGATTGAAAG   24
PAE_3          -GAATCTCAAAAAGAGGATTGAAAG   24
PAR_3          -GAATCTCAAAAAGAGGATTGAAAG   24
TTX_4          -GAATCTCAAAGAGAGGATTGAAAG   24
TPE_5          GTATCAACAACGAATGAGTTGAAAG   25
SMA_8          GTAACAACATAGAAAGAATTGAAAG   25
                       *       *       *  * *****
                       G.w....C....w..GrrTwGAAAG
```

Fig. 4.4: Conservation of crenarchaeal repeat sequences. 13 repeat sequences mapped in a multiple sequence alignment (ClustalW), conserved residues in red and ambiguity characters marked in yellow. MSE_4: *Metallosphaera sedula* CRISPR locus 4, SIS_3: *Sulfolobus islandicus L.S.2.15* CRISPR3, APE_3: *Aeropyrum pernix* CRISPR3, STO_8: *Sulfolobus tokodaii* CRISPR8, SAC_4: *Sulfolobus acidocaldarius* CRISPR4, SSO_3: *Sulfolobus solfataricus* CRISPR3, PCA_6: *Pyrobaculum calidifontis* CRISPR6, TNE_10: *Thermoproteus neutrophilus* CRISPR10, PAE_3: *Pyrobaculum aerophilum* CRISPR3, PAR_3: *Pyrobaculum arsenaticum* CRISPR3, TTX_4: *Thermoproteus tenax* CRISPR4, TPE_5: *Thermofilum pendens* CRISPR5, SMA_8: *Staphylothermus marinus* CRISPR8

All seven CRISPR loci in *P. furiosus* potentially produced small RNA transcripts (Hale *et al.*, 2008). In contrast, in *T. tenax* only the corresponding small RNA transcripts for five of seven CRISPR loci could be detected. For the CRISPR of group II, TTX_2 and TTX_3, no signals appeared on the blot. One explanation for these results could be the missing secondary structure of group II repeats, in comparison to the stem-loop structures of group I cluster, resulting in a low stability of generated group II small CRISPR RNAs (Fig. 3.5). Kunin and co-workers (2007) classified CRISPR by their repeat sequence similarity, defining 12 major groups. Overall, the repeats showed only minor or no dyad symmetry. Among the major groups, five archaeal groups were

defined, which were nearly all characterised as harbouring unstructured repeat secondary structures. This is in contrast to the predictions observed for the group I repeat sequences of *T. tenax* (Fig. 3.5). Also the secondary structure predictions of crenarchaeal repeats depicted in Fig 4.4 revelaed a similar stem-loop structure with minimum free energy values of averaged -3.39 kcal/mol (±1.84 kcal/mol; RNAfold). Therefore, it can be assumed that transcribed crenarchaeal CRISPR repeats form indeed remarkable secondary structures with lower stability, due to different thresholds chosen for significant secondary structures.

4.2 *cas* genes and their transcription

Four genes, always located near to a CRISPR array and only found in species containing CRISPR sequences, were identified in numerous prokaryotic genomes and therefore designated CRISPR-associated (*cas*) genes (Jansen *et al.*, 2002). These *cas* genes 1-4 are usually orientied head-to-tail as if they are cotranscribed and believed to be involved in the function of CRISPR.

4.2.1 Genomic context of *cas* genes in *Archaea*

Because of the constantly growing number of available prokaryotic genome sequence information, 41 new *cas* gene families were identified, located near to CRISPR arrays, in addition to the four previously described genes (Haft *et al.*, 2005). On the basis of genome comparisons, eight subtypes were defined in prokaryotes with clearly different sets of *cas* genes. Each subtype was named after a well-known representative comprising only one gene set, as numerous genomes have more than one subtype. The comparison of identified *cas* genes of *T. tenax* with these eight subtypes indicates that the *T. tenax* core gene cluster is very similar to the *A. pernix* subtype A (Fig. 3.6 and 4.5). Furthermore, in *T. tenax* a RAMP gene cluster could be defined which, is in parts homologous to the *M. tuberculosis* subtype M. The subset gene cluster shows no direct homology to the subtypes and is probably an inoperable relict. Degenerated *cas* gene clusters are a prevalent observation in prokaryotes, suggesting that these clusters are mobile adaptation modules useful for rapid adaptation to new ecological niches (Haft *et al.*, 2005). The *cas* genes are presumed to be involved in processes that include maintenance of repeat cluster, capture of new spacer elements, expansion and contraction of cluster, propagation of leader

sequences and repeat clusters within a genome, transfer of CRISPR and *cas* genes to new genomes and interaction of CRISPR/*cas* loci within the host cell (Makarova *et al.*, 2006). The hypotheses of the molecular functions of CRISPR/Cas systems especially in archaeal cells will be discussed later in detail (s. 4.3.1).

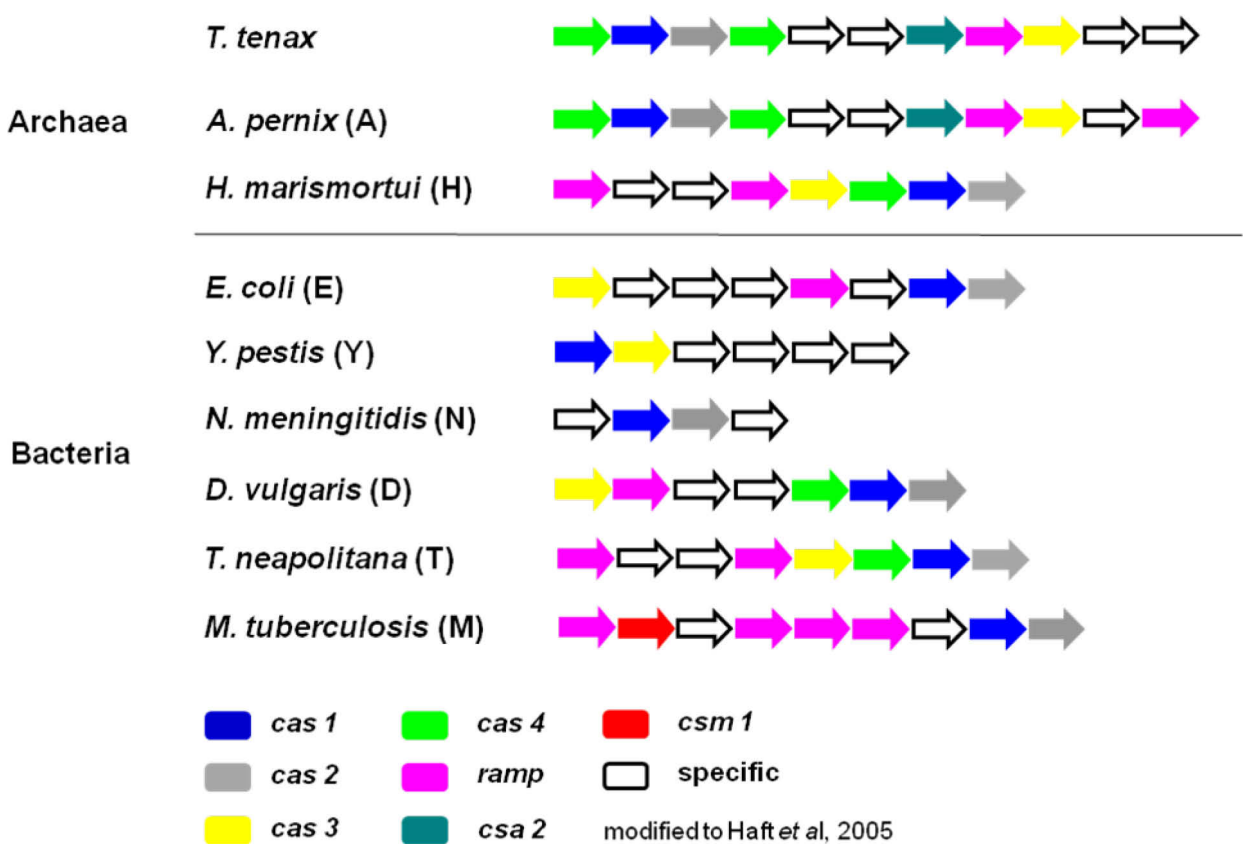


Fig. 4.5: Genomic context analyses of *cas* genes in prokaryotes. Core *cas* gene cluster of *T. tenax* in comparison to eight subtypes (A – Y) identifed of Haft and co-workers (2005) in prokaryotic genomes. Only selected *cas* genes (*cas1-4*, *ramps*, *csa2* and *csm1*) coloured, subtype specific genes marked by white arrows.

In accordance to the previous comparisons of 24 archaeal genomes, these genomes were analysed for their number and distribution of *cas* genes. Only genes were taken into account that were located nearby identified CRISPR array and showed significant homology to the 45 *cas* gene families (Tab. 4.2). In general, archaeal organisms are not only provided with numerous CRISPR, as shown before, but also with a higher number of *cas* genes in comparison to *Bacteria*. Hereby, the number of *cas* genes correlates with the optimal growth temperatures of the organism: *Archaea* with a higher optimal growth temperature comprehend generally more *cas* genes (Fig. 4.6). This tendency was already observed for the number of CRISPR and spacers, indicating a congruent evolution of CRISPR and *cas* elements (Fig. 4.1).

Tab. 4.2: Comparison of 24 archaeal organisms and their *cas* genes. Listed are number of all identified *cas* genes according to families of Haft *et al.*, 2005, number of *cas* gene cluster in the genome and optimal growth temperatures.

Organism	*cas* genes	Gene cluster	T_{opt}°C
Crenarchaeote			
Aeropyrum pernix	12	1	90
Staphylothermus marinus	15	1	90
Metallosphera sedula	22	3	65
Sulfolobus solfataricus	54	5	85
Sulfolobus tokodaii	30	3	75
Sulfolobus islandicus L.S.2.15	21	1	80
Thermofilum pendens	22	3	88
Pyrobaculum aerophilum	27	3	98
Pyrobaculum arsenaticum	22	2	95
Pyrobaculum calidifontis	29	2	90
Thermoproteus neutrophilus	32	3	85
Thermoproteus tenax	21	2	86
Euryarchaeote			
Archaeoglobus fulgidus	25	3	85
Haloarcula marismortui, pNG400	8	1	37
Methanosphaera stadtmanae	15	2	37
Methanocaldococcus jannaschii	19	2	85
Methanospirrilum hungatei	28	3	37
Methanococcoides burtonii	15	2	23
Methanosarcina acetivorans	14	2	40
Methanosarcina barkeri	16	3	37
Methanosarcina mazei	13	2	37
Pyrococcus furiosus	23	2	100
Thermococcus kodakaraensis	17	1	85
Picrophilus torridus	14	2	55

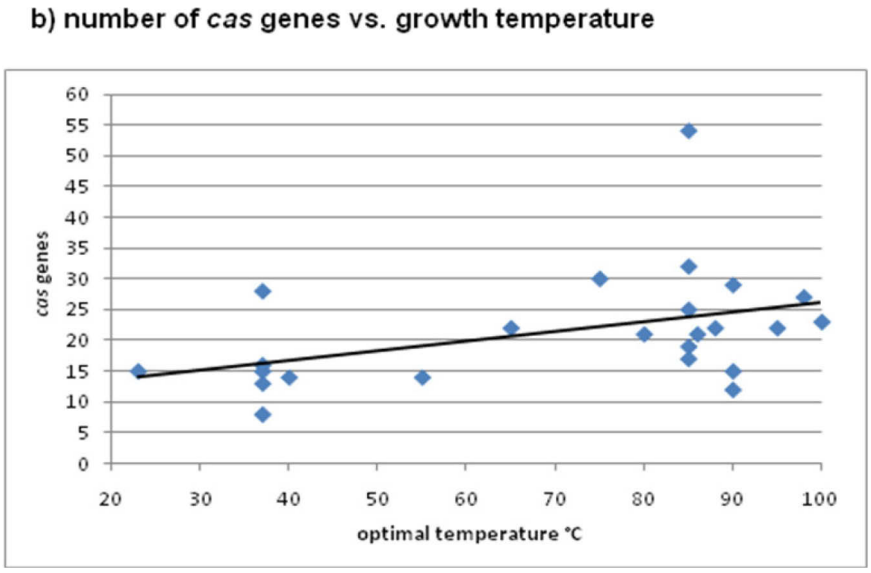


Fig. 4.6: Correlation of *cas* genes and optimal growth temperatures in 24 archaeal organisms. Genomes listed in Tab. 4.2 analysed for number of *cas* genes and optimal growth temperatures (source: DSMZ).

4.2.2 Transcription of *cas* genes

The basal archaeal transcription apparatus is more closely related to the core components of *Eucarya* than of *Bacteria* (Bell and Jackson, 1998). Features of that close relation are: *Archaea* contain a type II-like DNA dependent RNAP, which resembles the eucaryal nuclear RNAPs in subunit composition (Zillig *et al.*, 1979); *Archaea* comprehend homologs of several eucaryal transcription factors, including the TATA-box binding protein (TBP) and transcription initiation factor IIB (TFIIB; Bell and Jackson, 1998; Bell *et al.*, 2001). Transcription is regulated by sequence-specific transcription factors that prevent, stimulate or stabilise TBP, TFB and/or RNAP binding to the promoter region. Surprisingly, most of the so far identified transcriptional regulators are bacterial-like (Kyrpides and Ouzounis, 1999; Aravind and Koonin, 1999). Thus, archaeal transcriptional regulation represents a hybrid of bacterial and eucaryal features.

In *Archaea*, the majority of genes are organized in operon structures (50-70 %), containing on average three genes and subsequently producing polycistronic mRNA transcripts (Torarinsson *et al.*, 2005). The main sequence logos of the promoter are the box A motif (TATA-box: positions between -29 and -23 nt) and the B recognition element motif (BRE-site: 3 nt upstream from box A), interacting with TFB (Bell *et al.*, 1999). *Archaea* are not only producing polycistronic mRNAs, but also have typical Shine-Dalgarno (SD) motifs within transcripts. These are common mechanisms of

translational initiation shared with *Bacteria*, but the enzymatic machinery is far more complex. Numerous initiation factors, including a homologoue of the heteromeric protein complex eucaryal eIF2 were identified, representing once again the mosaic character of archaeal information processes (Keeling and Doolittle, 1995).
Analyses of archaeal genomes revealed some exceptional properties of transcription initiation and translation (Tolstrup *et al.*, 2000; Slupska *et al.*, 2001). A high percentage of leaderless transcripts, produced from single genes and first genes of operons were identified. Leaderless mRNAs are defined as transcripts missing typical SD motifs in front and the translational start codon is located right at the 5'-end or just a few nucleotides downstream (Londei, 2005). Leaderless mRNAs are, regardless of their source, universally translatable by bacterial, eucaryal and archaeal ribosomes. This indicates leaderless transcripts as a very ancient mechanism, still operative in *Archaea* (Grill *et al.*, 2000).

The *cas* genes from *T. tenax*, presented in this work, are a typical example for archaeal gene organisation and transcription. The core *cas* genes are organised in two operon structures, named *casa1* and *casa2*. In front of both operons TATA-boxes and BRE-sites could be identified and the functional genes within the operons showed overlapping start and stop codons. Furthermore, the polycistonic transcripts of both operons could be identified by RT-PCR analyses (Fig. 3.12). The transcript of *casa2* is clearly leaderless, as five of the six genes within the operon exhibit consensus SD motifs (GGAG or GGGG, at positions -7 to -4 nt of start codon, Appendix A4). This allows an efficient decoding of polycistronic messages in a long mRNA. Termination of transcription in *Archaea* is not strictly dependent on T tracts, but also on AT-motifs of eight nucleotides at the 3'-end. But occurrence of termination can happen at multiple adjacent sites, so that transcripts often have defined 5'-ends but multiple 3'-ends (Santangelo *et al.*, 2009). For *casa2* an eight nt long AT-motif is located 14 nt downstream of the last gene in the operon, indicating a putative termination site for *casa2* transcripts. The *casa1* transcript also seems to be leaderless, but although a polycistronic transcript could be detected, a gap of 89 bp is located between the genes *csa1* and *cas1*/*2*. But obviously this is not influencing the complete transcription of the operon. Consensus SD motifs GGAG or GGGG were located -12 to-15 nt of start codons. Termination motifs (two times 6 nt long AT-motif

at positions 18 and 46 downstream) were located downstream of the last gene in the operon, but not within the gap.

The analyses of *T. tenax cas* genes clearly showed that the transcription levels of *cas3* and *cas4* were affected by abiotic stress parameters in comparison to control cells (Fig. 3.9, 3.11), indicating that regulatory mechanisms control the transcription. The *cas3* gene showed a more than threefold increased and the *cas4* gene a twofold increased transcript level in cells treated 2 min with UV-light in comparison to the control cells. Furthermore, *cas3* mRNA levels of cells treated with 100 mM NaCl were more than tenfold increased and at 150 mM NaCl still fivefold increased in comparison to the control. As the studied genes *cas3* and *cas4* are located within operon structures producing polycistronic mRNAs, it can be assumed that the results are also true for *casa1* (*cas4* - *cas1*/*2* - *csa1*) and *casa2* (*csa5* - *csa2* - *cas5a* - *cas3* - *cas3hd* - *csa4*). The mRNA levels of the single gene *csa3* were only slightly affected (Plagens *et al.*, 2010a; in preparation). Different temperatures up to 91°C had generally no effect on the transcription of *cas* genes, but ongoing studies in our working group try to investigate the influence of temperatures above 91°C.
These results could be confirmed by other studies in *Archaea* using microarray technologies for identification of differences in transcript levels under stress conditions. *M. jannaschii* has an optimal growth temperature of 86°C and was heat shocked from 88 to 98°C (Booyaratanakornkit *et al.*, 2007). Up-regulated by greater than twofold were genes coding for the prefoldin α subunit, the thermosome subunit, small heat shock proteins and Csa2 (MJ0381). Exposure of *P. furiosus* to ionizing radiation at 2,500 Grays resulted in chromosome fragmentation, but within 2-4 h of incubation the chromosomes were reassembled (DiRuggiero *et al.*, 1997), indicating very potential DNA-repair systems. The harmful effects of ionizing radiation are the production of hydroxyl radicals, oxidative stress, lesions to DNA and cross-linking of proteins to DNA (Riley, 1994). Microarray data of *P. furiosus* exposed to ionizing radiation revealed that the polycistronic transcript of one *cas* operon were up to tenfold increased (PF0637-0643: *csa4* - *csa4* - *cas3hd* - *cas3* - *cas5a* - *csa2* – *csa5*). A second *cas* operon (PF1117-1122: *cas2* - *cas1* - *cas4* - *cas3* - *cas5* - *csa2*) showed no significant changes in gene expression levels (Willliams *et al.*, 2006). Noteworthy, the gene organisation of the first *cas* operon, regulated by gamma irradiation, is identical to *casa2*. UV-light stress (200 J/m^2) studied in *S. solfataricus* revealed that

cas genes were only slightly effected: 1.5 to 2 h after exposition *csa4*, *csm4*, *cmr2* were two- to threefold up-regulated (Götz *et al*., 2007). Not only the influence of abiotic stress parameters was studied, but also if a phage infection has an effect on *cas* gene transcription, supporting the hypothesis of a defense system against mobile genetic elements. Transcriptome analysis of *S. solfataricus*, infected with *Sulfolobus turreted icosahedral virus* (STIV), showed that transcription of viral genes started 8 h after infection leading to cell lysis after 32 h (Ortmann *et al.*, 2008). However, no significant transcription differences for *cas* genes were observed (*cas1*, approx. two-fold increased at 24 h postinfection). In summary, abiotic parameters significantly modulate the transcription levels of *cas* genes. The molecular function of the respective *cas* genes, especially of genes from *casa2* (*cas3*, *csa2*, *csa4*) remain uncertain, but indicate an involvement of the encoded proteins in the cellular response to harmful environmental changes.

4.2.3 Is Csa3 a transcriptional regulator for *cas* genes?

As mentioned above, transcription is regulated by sequence-specific transcription factors acting at the promoter region. Several transcriptional regulators have been characterised in *Archaea*, e.g. the metal dependent repressor 1 (MDR1) from *Archaeoglobus fulgidus*, a homolog of metal-dependent bacterial repressors. MDR1 represses transcription of its own gene and of an ABC-metal transporter encoded in the same operon, in a metal dependent manner (Bell *et al.*, 1999). A maltose-specific regulator (TrmB) for the trehalose/maltose transport operon from *Thermococcus litoralis* has also been characterized, inhibiting transcription counteracted by maltose and trehalose (Lee *et al.*, 2003). Examples for positive gene regulation are described for *Methanothermobacter thermautotrophicus*, where Tfx is proposed to activate expression of the *fmdECB* genes coding for molybdenum formylmethanofuran dehydrogenase (Hochheimer *et al.*, 1999).

The *T. tenax* Csa3 protein is a potential candidate for transcription regulation of *cas* gene operons, because *csa3* is located between the two gene clusters of *casa1* and *casa2* and contains a typical DNA-binding HTH-motif. The purification of the recombinant protein and the EMSA experiments of Csa3 with promoter regions confirmed that the protein binds unspecifically DNA (Fig. 3.17). Therefore, it was necessary to establish a protein purification protocol in which the DNA contamination is removed by PEI precipitation (Burgess, 1991; Hardy and Martin, 2008). The sequence of Csa3

not only showed significant similarity to other archaeal Csa3 proteins, but also to the putative TrmB protein of *Pyrobaculum aerophilum*. The alignment of selected archaeal homologs of Csa3 revealed two significant motifs (motif 1: TXGF(X)$_6$RA and motif 2: GGXR) without known function and the HTH-motif at a similar position in all sequences (Fig. 4.7), indicating a similar regulatory function in many crenarchaeal species.

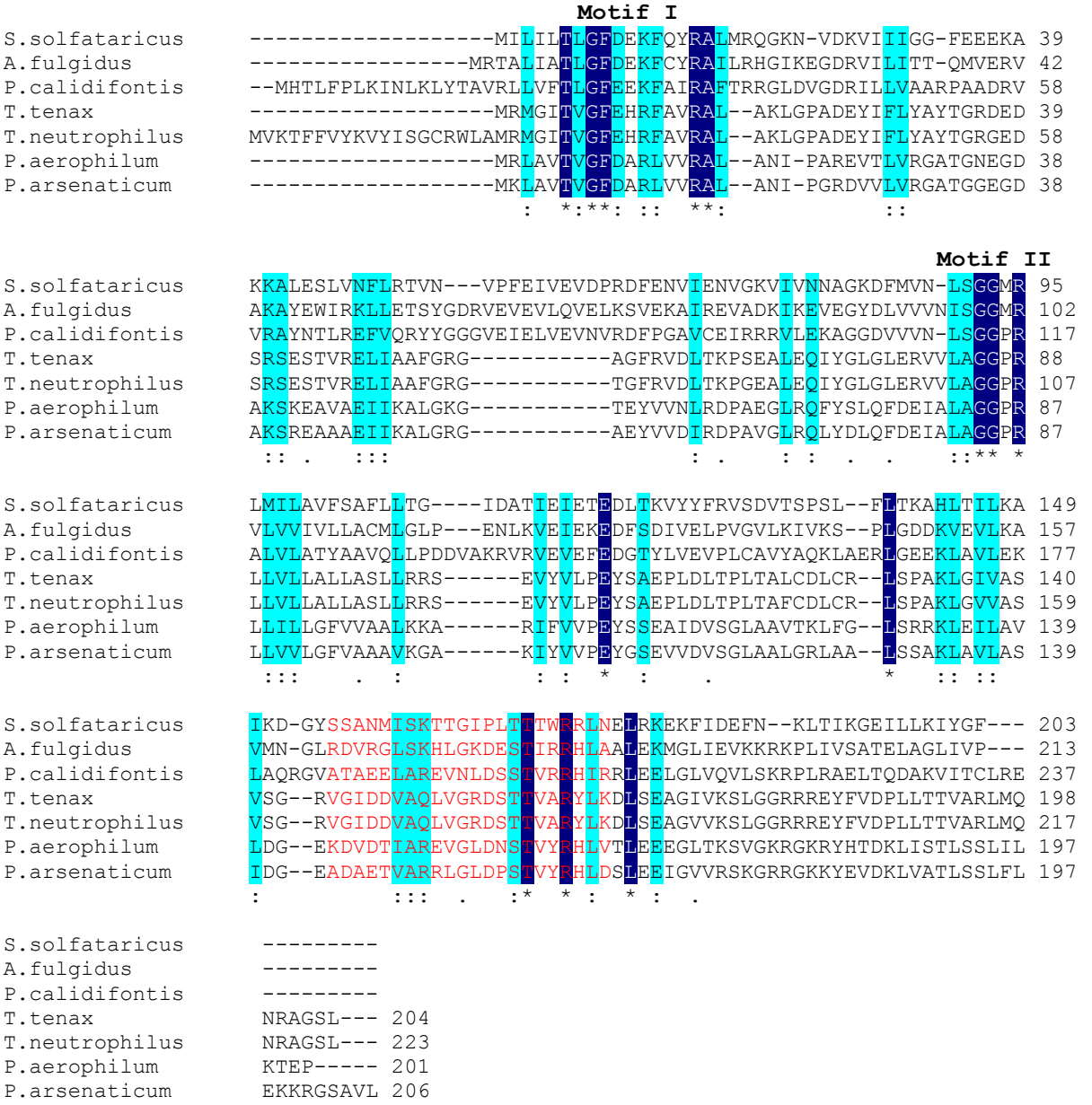


Fig. 4.7: Multiple sequence alignments of archaeal Csa3 homologs. The alignment was performed with CLUSTALW2. Conserved (*) and chemically similar (. and :) amino acid residues are marked. The putative DNA-binding (HTH motif) domain is marked in red (predictions were performed by NPS@); motif 1: TXGF(X)$_6$RA and motif 2: GGXR marked. Selected protein sequences: SSO_1444, AF_1869, Pcal_0266, TTX_1249, Tneu_1138, Pae_0202, Pars_1125.

Low molecular weight compounds (e.g. metals, sugars, nucleotides) are often responsible for the specific binding of transcriptional regulators. So, an intensive similarity search was conducted to define putative compounds for testing in future experiments. As mentioned previously, TrmB is involved in maltose-specific regulation of gene clusters in archaeal organisms (Lee *et al.*, 2003). TrmB and TrmB-like proteins have also the potential for binding trehalose or glucose at specific binding-sites (Lee *et al.*, 2007). As the *T. tenax* Csa3 showed high similartiy to an annotated TrmB homolog from *P. aerophilum*, the identified sugar-binding site was compared, but no similarity could be detected. Haft and co-workers (2005) defined Csa3 as distantly related to ArsR, which was also observed for *T. tenax* Csa3 in PSI-BLAST searches. Proteins of the ArsR family bind different metal ions, such as arsenate, cadmium or zinc, inducing a conformational change preventing the protein from binding DNA (Bairoch, 1993). Cyclic AMP (cAMP) receptor proteins (CRP) are global transcriptional regulators in *Bacteria*. In *E. coli* they control the activity of over 100 genes (McKay and Steitz, 1981). The cellular roles are diverse, including carbohydrate metabolism, development of competence for transformation, modulation of virulence gene expression and pathogenesis (Botsford and Harman, 1992). In *Thermus thermophilus* a CRP homolog (TTHA1437) controlled the transcription of a multitude of *cas* genes, such as *cas1-3*, *csm1-5* and *cse1-5e* (Shinkai *et al.*, 2007). Therefore, potential candidates for binding studies of Csa3 are maltose or trehalose, different metal ions or cAMP to confirm if Csa3 is a potential transcriptional regulator for *cas* genes in *T. tenax* and other *Archaea*.

4.3 The CRISPR/Cas system in prokaryotes

Recently, it became clear that CRISPR and the associated *cas* genes constitute a functional unit. The *cas* genes were presumed to be involved in processes that include maintenance of repeat cluster, integration of new spacer elements, expansion and contraction of cluster and interaction of CRISPR/*cas* loci within the host cell (Makarova *et al.*, 2006). One of the important roles of the CRISPR/Cas system is the defense of prokaryotes against mobile genetic elements. But an involvement in other cellular processes cannot be excluded. The following part will give an overview of working models for the functions of the CRISPR/Cas systems in prokaryotes and tries to integrate the results obtained for *T. tenax*.

4.3.1 The different roles of CRISPR/Cas systems

Some of the spacer sequences found in CRISPR matched fragments of extra-chromosomal elements, including bacteriophages and plasmid sequences (Bolotin *et al.*, 2005; Mojica *et al.*, 2005). Furthermore, *Sulfolobus islandicus rod-shaped virus 1* that penetrate the cell are unable to infect *S. solfataricus*, when CRISPR array contain spacers similar to the phage genome (She *et al.*, 2001) or the plasmid pNOB8 transferred into *S. solfataricus* is not integrated or maintained as a replicon (Schleper *et al.*, 1995). This findings led to the hypothesis that the CRISPR/Cas system might be a novel defense system against invading alien nucleic acids (Makarova *et al.*, 2006).

All types of cellular life are potential targets for invasion by mobile genetic elements, such as viruses, plasmids and transposons. To establish a balance between uptake and integration on the one hand (e.g. competence, conjugation, recombination) and stabilisation of the genome integrity on the other hand, a variety of defense systems has been evolved that limit horizontal gene transfer (HGT; van der Oost *et al.*, 2009). In mammals, the adaptive immune system targets viruses and pathogens by antibodies secreted by B cells and viral epitopes on infected cells by antigen receptors of cytotoxic T cells (Guidotti and Chisari, 2001). The small interfering RNA (siRNA) system relies on the recognition of RNA genomes of retro-viruses and retro-transposons. After detection, dsRNA is processed by Dicer endonucleases and short duplexes are bound and unwound by the argonaute subunit of the RISC complex. The remaining guide strand is responsible for specific target recognition by base-pairing and degradation of foreign mRNA (Jinek and Doudna, 2009). Many anti-virus systems have also been described for prokaryotes. Restriction enzymes recognize and degrade DNA without methylation and with a non-native methylation pattern at specific sites. In addition, several other defense systems have been identified operating at different levels of the viral lytic cycle, e.g. impairing of viral cell adsorption or blocking of DNA injection (Sturino and Klaenhammer, 2006).

The CRISPR/Cas system operates as a novel defense system that is able to protect a host cell against invading foreign nucleic acid (Makarova *et al.*, 2006). The first experimental evidence was acchieved in *S. thermophilus*. Phage-resistant mutants had inserted some additional spacers at the leader sequence site of a CRISPR loci after infection with two viruses. The acquired spacers are identical to sequences of the phage genome. As shown, a 100 % identity of the spacers is necessary, as

single-nucleotide polymorphisms of 1 to 15 nt within the spacer abolished the resistance of the organism. Inactivation of a *cas* gene (*csn5*) results also in loss of phage resistance (Barrangou *et al.*, 2007). Furthermore, the CRISPR/Cas system prevents conjugation and plasmid transformation within the organism. One CRISPR spacer with high similarity to the *nickase* gene present in its conjugative plasmids was observed. As shown, the CRISPR/Cas system limits the horizontal gene transfer in *S. epidermidis* by targeting plasmid DNA (Marraffini and Sontheimer, 2008). Thus, the CRISPR/Cas system is not only limited to phage defense, but has a more general role in prevention of HGT and maintenance of genetic integrity.

An interesting observation was made by Zegans and co-workers (2009), indicating also a regulatory function of the CRISPR/Cas system. The lysogenic infection of *Pseudomonas aeruginosa* with the bacteriophage DMS3 inhibits biofilm formation and swarming motility. Mutation or deletion of five *cas* genes and a CRISPR region restored biofilm formation and swarming motility, but spacer sequences similar to phage sequences did not prevent the infection and integration. Therefore, the CRISPR/Cas system modulates lysogeny by regulatory effects within the cell for limiting bacteriophage dissemination in bacterial communities. Also the detection of spacer sequences matching the own chromsome such as in *T. tenax* (Tab. 3.2) or in *P. aerophilum* (Mojica *et al.*, 2005) suggests a regulatory role.

Furthermore, it is hypothesised that the CRISPR/Cas system is involved in other cellular processes: i) chromosomal segregation (Mojica *et al.*, 1995; Charlesbois *et al.*, 1998); ii) homologous recombination, because of large scale DNA rearrangements in *Thermotoga*, due to the location of CRISPR loci at DNA joints that connect shuffeled chromosomal segments (DeBoy *et al.*, 2006); iii) DNA repair and thermal stability (Makarova *et al.*, 2002). Especially the last point is very interesting under the aspect of adaptation to high temperatures, short-wave radiation or ions. Thermophilic organisms contain more CRISPR loci, spacers and *cas* genes and the transcription of some *cas* genes of *T. tenax* and other *Archaea* clearly reacts on abiotic stress parameters, such as UV-light, ionizing radiation, salt or temperature (s. 3.2.2). Self-transmissable plasmids of *S. islandicus* are most stably maintained, when they incorporate a CRISPR array (Greve *et al.*, 2004). An *E. coli* strain adapted for 2,000 generations to 41.5°C resulted in duplication and recombination of CRISPR loci and this reorganisation is coincident with the increase in fitness of the organism (Riehle *et al.*, 2001). One reason for that may be hairpin structures in the genome

that lend to greater resilience under harsh conditions (Fadiel *et al.*, 2003). In particular, the thermophilic archaeal species were predicted to form a strong bending of 72° at the DNA repeat sequences that might result in a regular secondary structure (Jansen *et al.*, 2002).

Taken together, the results of the present work lead to the suggestion that the role of the CRISPR/Cas system in prokaryotes is complex: CRISPR and *cas* genes have the potential to be differentially regulated and co-opted for new regulatory functions. This can be useful in the rapid adaptation of new ecological niches (Haft *et al.*, 2005; Sorek *et al.*, 2008). It is undeniable that the CRISPR/Cas system is operating as a defense system against mobile genetic elements, presumably also in *T. tenax* due to identified spacer sequences similar to archaeal phages. As spacers are also homologous to the sequences of one´s own chromosome, an additional regulatory or interfering function within the genome is also likely. Moreover, a stabilising function of the genome structure cannot be excluded.

The CRISPR/Cas system in prokaryotes has the potential for numerous future applications in medicine and biotechnology, e.g. strain typing of pathogenic organisms, engineered virus defence or gene silencing.

Based on the observation that CRISPR loci are the most rapidly evolving structures in the genome, a spacer-oligotyping (also termed spoligotyping) method for strain detection was developed (Groenen *et al.*, 1993; Kamerbeek *et al.*, 1997). This has become the standard method for genotyping *M. tuberculosis* strains as part of ongoing efforts to control tuberculosis outbreaks (Brudey *et al.*, 2006), but was also established for the pathogenic prokaryotes *Campylobacter jejuni* and *Yersinia pestis* (Price *et al.*, 2007; Cui *et al.*, 2008). The CRISPR/Cas system might also offer a partial solution to the problem of culture losses that are caused by phages in the dairy or wine industry (Sturino and Klaenhammer, 2006). Artificially added spacers that are derived from conserved regions of known phages to a CRISPR array can support the immunity of industrial *Bacteria* against viruses (Sorek *et al.*, 2008). Furthermore, a modulated CRISPR system can help to prevent spreading of antibiotic resistance and virulence factors by blocking the horizontal gene transfer of pathogens (Maraffini and Sontheimer, 2008) or allow selective gene knockdown of endogenous genes (Makarova *et al.*, 2006).

4.3.2 Function of Cas proteins

Comparative genome analyses reveal that there is a huge diversity in the genetic composition of *cas* genes in the prokaryotic CRISPR/Cas systems (Haft *et al.*, 2005). The respective encoded proteins belong to a wide range of different protein families (s. 3.2.1). The distinct Cas proteins are probably involved at different functional stages of the CRISPR/Cas mechanism: i) CRISPR expression and processing to small RNAs; ii) CRISPR interference, the binding or degeneration of the target; iii) CRISPR adaptation, integration of new spacers (van der Oost *et al.*, 2009).

In *T. tenax* two Cas protein complexes could be identified within this work, named CasA1 and CasA2, incorporating members of the core Cas proteins (Cas1-4). The three proteins of the CasA1 encoded by the operon *casa1* (Cas4, Cas1/2 and Csa1) strongly interact in the refolding process. The reconstituted CasA1 complex showed RNase activity at protein concentrations of 12.5 - 17 ng/µl (Fig. 3.21; Plagens *et al.*, 2010b, in preparation).

In the following, focus lies on the experimentally confirmed results of members of the CasA1 complex in other organisms. Cas1 is an universal marker of the CRISPR/Cas system, represented nearly in all genomes (Makarova *et al.*, 2006) and has a strong basic isoelectric point (pI) of 9 to 10 (Jansen *et al.*, 2002), in *T. tenax* even larger than 11. Cas1 of the bacterium *Pseudomonas aeruginosa* showed endonuclease activity specific for DNA in a metal-dependent manner, generating fragments of ~80 bp (Wiedenheft *et al.*, 2009). In contrast, Cas1 of Archaeum *S. solfataricus* bound ssDNA, dsDNA, ssRNA and dsRNA with high affinity in a non-specific manner, but no nuclease activity was detected (Han *et al.*, 2009). One reason for different activities of Cas proteins is potentially caused by the diversity of Cas proteins found in prokaryotes that may reflect an involvement in significantly different mechanisms of the CRISPR/Cas system (Hale *et al.*, 2009). Cas2 of *S. solfataricus* was characterised as a Mg-dependent endoribonucleases, specific only for ssRNA. No specificity was observed for CRISPR transcripts and the preferential cleaveage site was located within U-rich regions. The structure of Cas2 revelaed a ferredoxin-like fold, present in numerous RNA-binding proteins (Beloglazova *et al.*, 2008). Cas4 and Csa1 are not yet enzymatically characterised, but both proteins belong to a restriction endonuclease-like superfamily, defined by a relatively conserved PD-(D/E)XK-motif and are functional similar to RecB exonucleases (Fig. 4.8). The distinctive motif

presumably marks the active site, contributing to cleaveage of the nucleic acid phosphodiester bond (Kinch *et al.*, 2005).

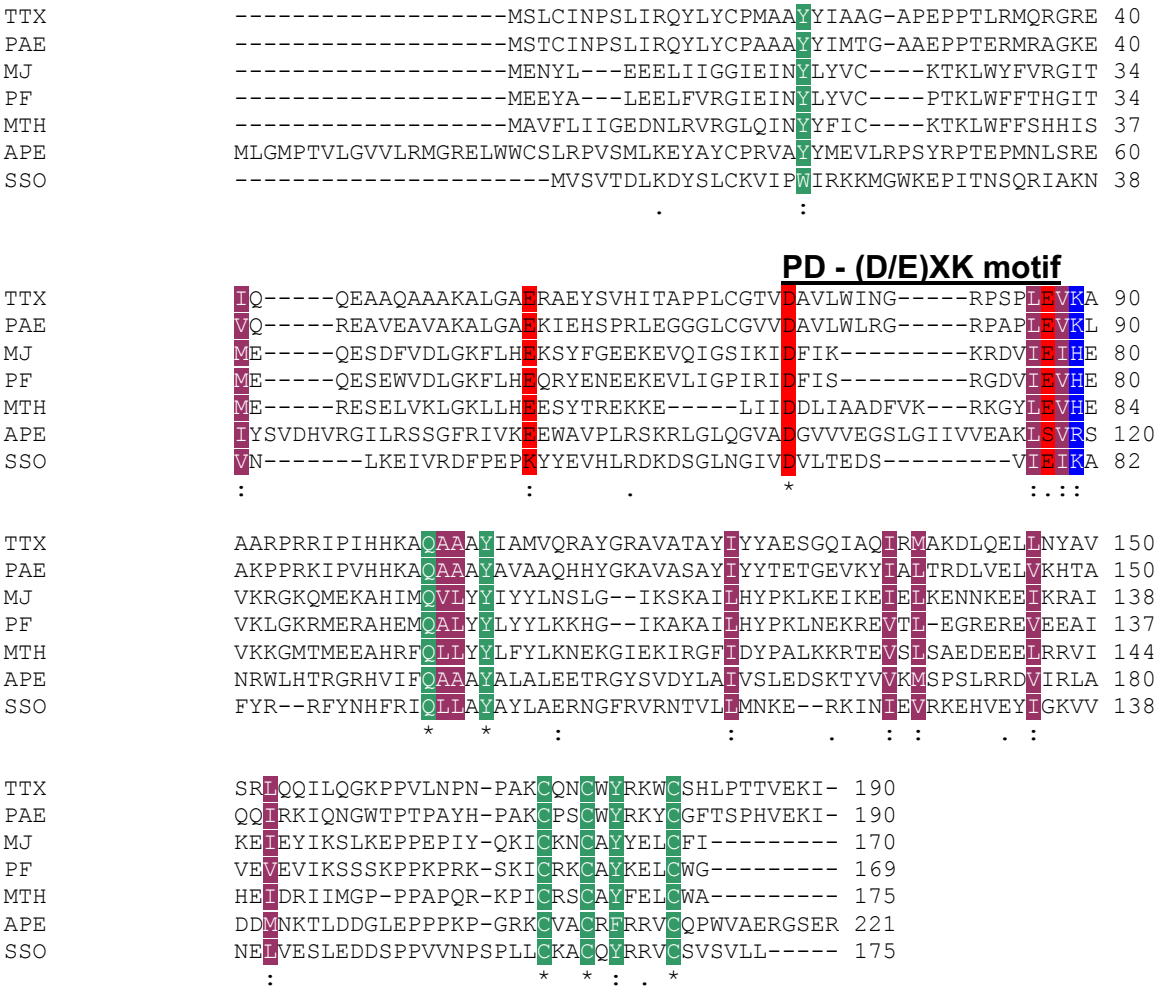


Fig. 4.8: Multiple sequence alignment of Cas4. Conserved residues and PD-(D/E)XK-motif highlighted: basic (blue), acidic (red), neutral (purple), polar (green). Sequences aligned with CLUSTALW2: *T. tenax* (TTX_1245), *P. aerophilum* (PAE_0198), *M. jannaschii* (MJ_0377), *P. furiosus* (PF_1119), *M. thermautotrophicus* (MTH_1085), *A. pernix* (APE_1239) and *S. solfataricus* (SSO_1449).

Nuclease activity of the *T. tenax* CasA1 complex facilitated through one or all its components seem to be confident. Further experiments concerning the substrate specificity (RNA vs. DNA) are required to characterise the CasA1 complex of *T. tenax* in more detail. The different results obtained from other organisms, especially for the Cas1 protein, is most likely based on the fact that only single components of the complex were analysed, and not the whole complex. The established reconstitution assay for CasA1 has the advantage for enabling defined mutagenesis approaches gaining insight into the roles played by individual components (Werner and Weinzierl, 2002).

The CasA2 complex of *T. tenax* is composed of the proteins Csa5, Csa2, Cas5a, Cas3, Cas3HD and Csa4. Similar to CasA1, all proteins encoded by the operon *casa2* strongly interact in the refolding process with most efficiency in the presence of all six proteins. Remarkable, a supporting effect of RNA in the reconstitution of CasA2 was determined as total *T. tenax* RNA was added. The RNA binding activity was also demonstrated by performing EMSA studies (Fig. 3.27).

Experimental results of members of the CasA2 complex in other organisms revealed a general interaction with nucleic acids. In *S. solfataricus* an interaction of Csa2 with SSB (single-stranded DNA-binding protein) was identified via pull-down experiments. SSBs play a central role in DNA replication, recombination and repair by localisation of DNA damages and DNA unfolding for recruitment of repair proteins (Cubeddu and White, 2005). It is noteworthy, that we observed DNA-binding activity for Csa2. After purification of the recombinant protein, an unspecific binding to *E. coli* DNA was observed comparable to the purification of the transcriptional regulator Csa3.

The Cas3 protein is a clear member of the DEAD/DEAH box helicases and all motifs for binding of substrates (Mg, ATP and RNA) and helicase activity could be identified in *T. tenax* Cas3 (Fig. 4.9).

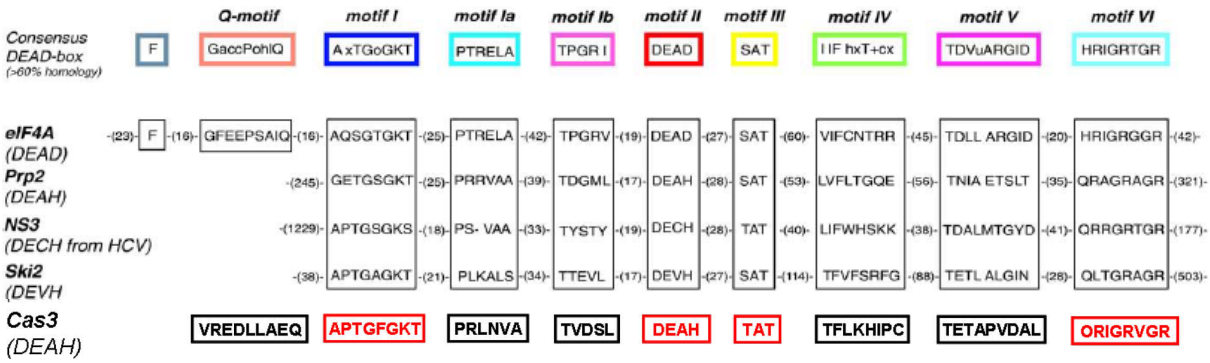


Fig. 4.9: Comparison of *T. tenax* Cas3 to DEAD/DEAH box helicase motifs. Different members of helicase family (eIF4A, *S. cerevisiae*; Prp2, *S. cerevisiae*; NS3, hepatitis C virus; Ski2, *S. cerevisiae*, Cas3, *T. tenax*) compared to consensus motifs. Q-motif: adenine recognition; motif I: Walker A motif, binding of Mg and ATP; motif Ia/b: RNA binding; motif II: Walker B motif, ATPase activity; motif III: ATPase and helicase domains linked; motif IV/V: RNA binding; motif VI: ATPase and RNA binding. Figure modified to Cordin *et al.*, 2006.

The Cas3HD protein contains a typical HD-motif, characteristic for the superfamily of metal-dependent phosphohydrolases (Yakunin *et al.*, 2004). The recombinant

Cas3HD protein from *S. solfataricus* was expressed in *E. coli*, but resulted in an insoluble form. The fusion with a soluble esterase domain enabled the solublity of the protein. This Cas3HD fusion protein degrades dsDNA and dsRNA unspecifically (Han and Krauss, 2009). Taken together, we suggest that the proteins of the CasA2 complex strongly interact with each other and thus it is important to study the function of the whole complex. CasA2 from *T. tenax* is a good candidate for further studies, since the complex is also present in other archaeal organisms, probably executing a similar function in these cells (Fig. 4.10).

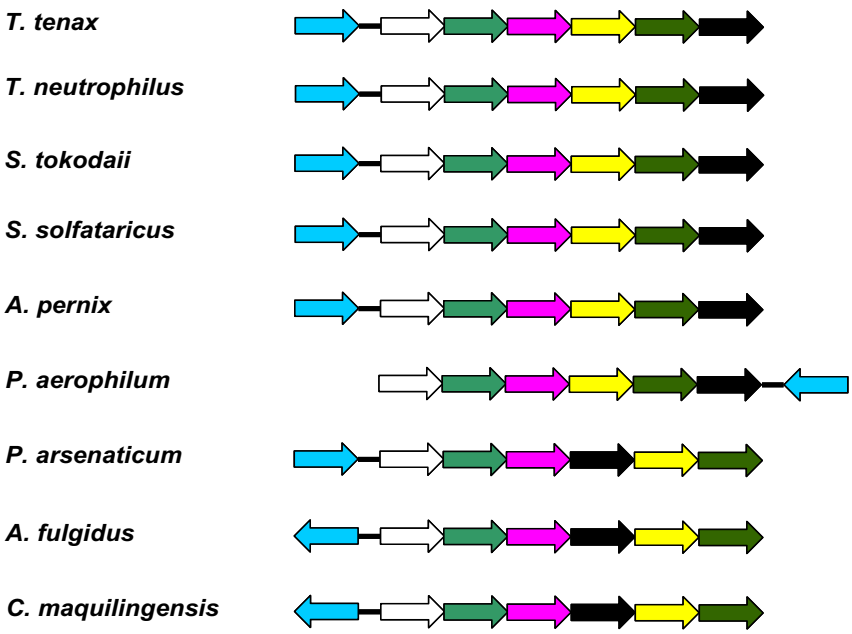


Fig. 4.10: Gene organisation of *casa2* in different archaeal organisms. Identical colours account for homolog genes: Csa3 (light blue), Csa5 (white), Csa2 (green), Cas5a (pink), Cas3 (yellow), Cas3HD (dark green), Csa4 (black). *T. tenax*: TTX_1249-55; *T. neutrophilus*: Tneu_1132-38; *S. tokodaii*: STS265-ST2641; *S. solfataricus*: SSO1438-44; *A. pernix*: APE1230-1237.1; *P. aerophilum*: PAE0202-0211; *P. arsenaticum*: Pars_1125-31; *A. fulgidus*: AF1869-75; *C. maquilingensis*: Cmaq_1520-26.

Several Cas proteins were characterised in different organisms, however, there are no homologs present in the *T. tenax* CRISPR/Cas system. In *E. coli*, protein complexes were identified composed of Cse1-5, termed Cascade complex (CRISPR-associated complex for antiviral defense). This complex cleaves long CRISPR transcripts specifically within the repeat sequence into small fragments of 57 nt length. The Cascade bound CRISPR RNA serves subsequently as a guide to direct the complex to viral nucleic acids to mediate an antiviral response (Brouns *et al.*, 2008). *P. furiosus* Cas6, a member of the RAMP superfamily, is an endoribonuclease

specific for the repeat sequences, generating a spacer unit with an 8 nt repeat-tag (Carte *et al.*, 2008). The sizes of these small RNAs are similar to the pattern observed in Northern blot analyses of total RNA in *P. furiosus* (s. 4.1.4). The produced CRISPR RNAs are subsequently bound to a protein complex composed of Cmr1-6 (RAMP module) and this complex cleaves complementary to small CRISPR RNA targets of ssRNAs at specific sites (Hale *et al.*, 2009).

4.3.3 The CRISPR/Cas system of *T. tenax* – A working model

In the following, I summarize what we know and suggest about the CRISPR/Cas system of *T. tenax* regarding: i) CRISPR expression; ii) CRISPR interference and iii) CRISPR adaptation (Fig. 4.11).

i) CRISPR expression

In *T. tenax* CRISPR loci are transcribed into RNA and step-wise processed to small CRISPR RNAs with a length of ~130, ~110, ~70 and ~50 nt. The transcription start site is located in front of the first repeat within the leader sequence distinguished by typical BRE-sites and TATA-boxes that facilitate the binding and operating of the basal transcription machinery. As known from other prokaryotic organisms, the processing of CRISPR transcripts is an enzymatically catalysed procedure. Reasonable candidates in *T. tenax* are particularly the CasA1 complex due to the detected ribonuclease activity, but maybe also another non identified protein. Proteins detected in other organisms for processing CRISPR transcripts are the Cascade complex (*E. coli*, Brouns *et al.*, 2008) and Cas6 (*P. furiosus*, Carte *et al.*, 2008). But, the genes encoding the Cascade complex are not present in *T. tenax* and Cas6, typically located next to the *cas* core genes, is identifed only at a fractional *cas* gene cluster. As the CRISPR transcription seems to be constitutive in *T. tenax*, it is therefore likely that the transcription of the involved *cas* genes is regulated. The transcription level of *cas* genes are modulated by abiotic stress parameters, such as UV-light or high ionic strength. The DNA-binding protein Csa3 located between the two *cas* gene operons *casa1* and *casa2* is a potential candidate for transcriptional regulation of *cas* genes, but it is unclear so far which signal, i.e. low molecular weight compounds, triggers the specific binding of the regulator to the promoter region and how the connection between environmental stress response and regulation of *cas* genes is implemented.

ii) CRISPR interference

Experiments showed that the CRISPR/Cas system guide antiviral defense and restrict horizontal gene transfer (Barrangou *et al.*, 2007; Marraffini and Sontheimer, 2008). However, no clear ideas about the interference reactions exists, which could result in a blocking or a degeneration of targets. A prerequisite for an interfering reaction is a template or guiding strand in order to allow a complementary binding of small CRISPR RNAs and protein complexes to its target. In *T. tenax* the CasA2 complex is a potential candidate for accomplishing such an interference reaction. Reasons for that are: i) strong interaction with RNA, detected in the reconstituion assays of the CasA2 complex and in EMSA studies, and ii) defined RNA and DNA binding domains of components of CasA2 complex. Further studies are necessary to identify the real target (RNA and/or DNA) and to describe the reaction mechanisms (blocking and/or degeneration) of CasA2 complex.

Potential targets of the interference reaction are assignable by sequence similarity of the spacers. As spacers are not only similar to foreign genetic sequences, such as viruses, phages or transposons, but also to own chromosomomal sequences, a regulatory function is possible, in addition to antiviral defense. Moreover, a stabilising function of the genome, of plasmids or other cell components is likely: i) the presence of CRISPR loci within the chromosome has a stabilising effect on the DNA structure and the number of CRISPR loci and *cas* genes correlates with an increased optimal growth temperature; ii) *cas* genes are modulated by environmental factors and the encoded proteins can have a stabilising or repairing impact on nucleic acids in the cell. Thus, we propose that the CRISPR/Cas system is not only operating as a defense system against mobile genetic elements, but has a multifunctional impact on the *T. tenax* cell (Plagens *et al.*, 2010b; in preparation)

iii) CRISPR adaptation

CRISPR are the most rapidly evolving elements in the genome (Groenen *et al.*, 1993). Closely related species that are to 99 % identical at DNA level, differ remarkably in their CRISPR composition (Pourcel *et al.*, 2005; Bolotin *et al.*, 2005). The locus evolves via polarized addition of novel spacers at the leader site and internal deletion, probably as a result of spontaneous recombination (Horvath *et al.*, 2008; Deveau *et al.*, 2008). There is evidence for acquisition of CRISPR loci and *cas* genes into the genome via HGT, as numerous CRISPR loci and *cas* genes are not

located on bacterial chromosomes, but were found on plasmids (Godde and Bickerton, 2006). The molecular mechanisms behind the addition of new spacers and the evolution of CRISPR loci are totaly unclear. The CRISPR/Cas system is a dynamic and inheritable system in prokaryotes with varying impact on the cells of different organisms. However, many aspects of the CRISPR/Cas system in *T. tenax* and in all prokaryotes remain unclear and require further experiments.

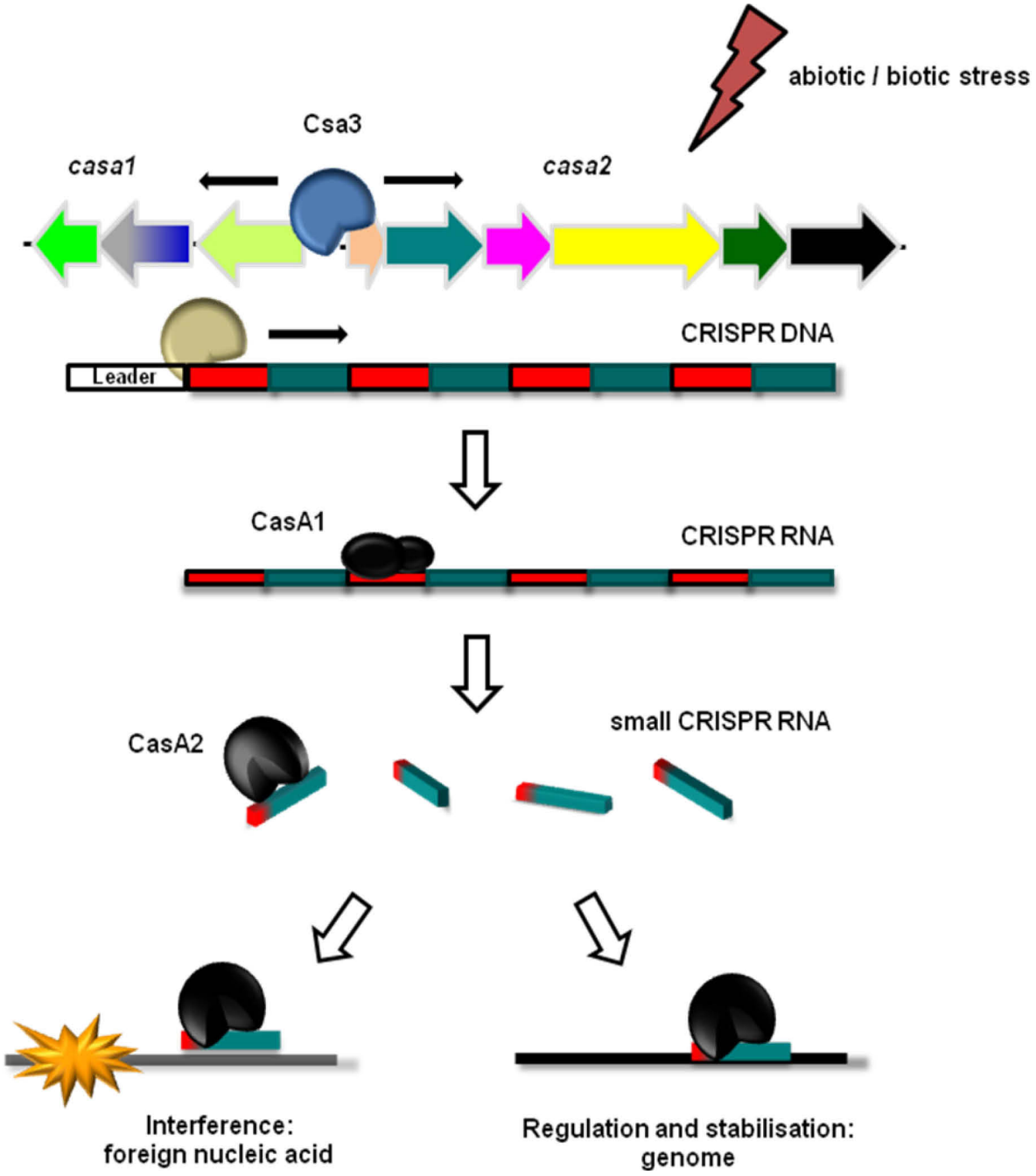


Fig. 4.11: Working model for the function of the CRISPR/Cas system in *T. tenax*

5. SUMMARY

Clustered regularly interspaced short palindromic repeats (CRISPR) found in prokaryotes are non-contiguous direct repeats with a length of 24-48 nt. The sequence repeats are weakly palindromic at the 5'- and 3'-termini and separated by variable spacer sequences of similar size (Jansen *et al.*, 2002). CRISPR loci are flanked on one side by an AT-rich leader sequence of 200-350 bp length. CRISPR sequences are widespread in the two prokaryotic domains; they were identifed in most archaeal genomes, in 40 % of the bacterial genomes and on some plasmids. A group of *cas* genes is always located near to a CRISPR locus and the encoded proteins are the essential actors for function and assembly of CRISPR. Recently, it could be demonstrated that the CRISPR/Cas system operates as a novel defense system, which is able to protect prokaryotic cells against mobile genetic elements, such as phages or plasmids. However, the role in prokaryotic genomes and the mechanisms that underlie CRISPR function are mostly uncharacterised, especially in *Archaea*. Therefore, the function of the archaeal CRISPR/Cas system is studied in more detail using the hyperthermophilic organism *Thermoproteus tenax* as an example.

In the course of this work seven CRISPR loci could be identified in the genome of *T. tenax*. The spacer sequences showed significant similarity not only to archaeal phages, but also to genes of the *T. tenax* genome. Northern blot analyses of small RNA species prepared from *T. tenax* cells with spacer probes from all seven CRISPR arrays, revealed transcript length of ~130 nt, ~110 nt, ~70 nt and ~50 nt, suggesting that large CRISPR RNA transcripts are stepwise endo- and exonucleolytically processed. The repeat sequence within the transcript has probably two functions: i) stabilising the CRISPR transcript by the formation of hairpin-structures and ii) binding motif for nuclease reaction. The transcription start site is located in front of the first repeat within the leader sequence distinguished by typical BRE-sites and TATA-boxes. The formation of small CRISPR RNAs is not regulated by abiotic stress factors, but the product of constitutive transcription in *T. tenax*.

Over 20 *cas* genes were identified in the genome of *T. tenax* and their organisation is similar to other crenarchaeal organisms. The core *cas* genes are organised in two operon structures, named *casa1* (*cas4*, *cas1*/*2* and *csa1*) and *casa2* (*csa5*, *csa2*, *cas5a*, *cas3*, *cas3hd* and *csa4*). In front of both operons TATA-boxes and BRE-sites

could be identified and the functional genes within the operons showed overlapping start and stop codons. Furthermore, the polycistronic transcripts of both operons could be identified by RT-PCR analyses. The polycistronic transcripts of *casa1* and *casa2* are clearly leaderless, whereas for all genes within the operons consensus SD motifs could be identified, which would allow an efficient translation of the genes within the mRNAs. Abiotic stress parameters, such as UV-light or high ionic strength, modulate the transcription of *cas* genes. The *cas3* gene showed a more than three-fold increased and the *cas4* gene a twofold increased transcript level in cells treated 2 min with UV-light in comparison to the control cells. Furthermore, *cas3* mRNA levels of cells treated with 100 mM NaCl were more than tenfold increased in comparison to the control.

Protein Csa3 can be considered as a good candidate for a transcription regulator, whose coding gene is located in between the two *cas* operons. This protein is structurally characterised by a typical HTH-motif and possesses high affinity to DNA. However, the conditions for a required specific binding to the promoter regions of *casa1* and *casa2* has still to be found. In future studies, the influence of low molecular weight compounds (e.g. trehalose, metal ions or cAMP), which could facilitate a specific DNA-binding behaviour, will be examined.

The functional characterisation of single Cas proteins encoded by the operons *casa1* and *casa2* was complicated by the formation of inclusion bodies of the recombinant proteins in *E. coli*. For *casa1*, reconstitution experiments revealed that the three proteins encoded by the operon strongly interacted in the refolding process. The resulting tripartite CasA1 protein complex showed ribonuclease activity with ssRNA CRISPR transcripts as a target at protein concentrations of 12.5 - 17 ng/µl. Also four of the six proteins encoded by the operon *casa2* could be expressed only in an insoluble form. For the other two proteins (Csa5 and Csa2) “normal” purification protocols could be established. Also in that case, the reconstitution of the multipartite complex (“CasA2”) is favoured in the presence of all proteins encoded by the *cas* gene operon. Remarkably, addition of *T. tenax* RNA supports the reconstitution of the CasA2 complex significantly. The RNA binding capacity could also be documented in EMSA studies.

The studies of the CRISPR/Cas system of *T. tenax* clearly showed that the processing of CRISPR transcripts is an enzymatically catalysed reaction. A relevant component in *T. tenax* is certainly the CasA1 complex due to its ribonuclease activity.

The CasA2 complex may be a candidate for accomplishing interference reactions, due to its strong interaction with RNA, detected in the reconstituion assays and supported by the presence of defined RNA and DNA binding domains of components.

The present data confirm the common hypothesis that the CRISPR/Cas system of prokaryotic cells represent a defense system against mobile genetic elements, considering the observation that some spacer sequences show clear sequence similarity to known viruses, plasmids or transposons. But there are also hints to additional functions: i) Regulatory functions may also be possible, since spacers of the *T. tenax* CRISPR are not only similar to foreign genetic sequences, but also to the organism's own chromosomomal sequences. ii) Moreover, functions of the CRISPR/Cas systems for stabilising the genome structure or nucleic acids are also likely, as the number of CRISPR loci and *cas* genes correlates with an increased optimal growth temperature and *cas* genes are modulated by environmental factors.

6. LITERATURE

Ahmed H, Tjaden B, Hensel R and Siebers B (2004) Embden-Meyerhof-Parnas and Entner-Doudoroff pathways in *Thermoproteus tenax*: metabolic parallelism or specific adaptation? *Biochemical Society Trans.*, 32(2), 303-304

Ahmed H, Ettema TJ, Tjaden B, Geerling AC, van der Oost J and Siebers B (2005) The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic Archaea: a reevaluation. *Biochem. J.* 390(2), 529-540

Ahn DG, Kim SI, Rhee JK, Kim KP, Pan JG and Oh JW (2006) TTSV1, a new virus-like particle isolated from the hyperthermophilic Crenarchaeota *Thermoproteus tenax*. *Virology* 351(2), 280-290

Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.* 10(2), 134-139

Allers T and Mevarech M (2005) Archaeal genetics – the third way. *Nat. Rev. Genet.* 6(1), 58-73

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215(3), 403-410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17), 3389-3402

Anderson I, Rodriguez J, Susanti D, Porat I, Reich C, Ulrich LE, Elkins JG, Mavromatis K, Lykidis A, Kim E, Thompson LS, Nolan M, Land M, Copeland A, Lapidus A, Lucas S, Detter C, Zhulin IB, Olsen GJ, Whitman W, Mukhopadhyay B, Bristow J and Kyrpides N (2008) Genome sequence of *Thermofilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction. *J. Bacteriol*. 190(8), 2957-2965

Aravin AA, Hannon GJ and Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761-764

Aravind L & Koonin EV (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.* 27(23), 4658-4670

Bairoch A (1993) A possible mechanism for metal-ion induced DNA-protein dissociation in a family of prokaryotic transcriptional regulators. *Nucleic Acids Res*. 21(10), 2515

Bankier AT, Weston KM and Barrell BG (1987) Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Meth. Enzymol.* 155, 51-93

Barlow JJ, Mathias AP, Williamson R and Gammack DB (1963) A simple method for the quantitative isolation of undegraded high molecular weight ribonucleic acid. *Biochem. Biophys. Res. Commun*. 13, 61-66

Barns SM, Delwiche CF, Palmer JD and Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA* 93(17), 9188-9193

Barrangou R, Fermaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA and Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes, *Science* 315, 1709-1712

Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, Koonin EV, Edwards AM, Savchenko A and Yakunin AF (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 283(29), 20361-20371

Bell SD and Jackson SP (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* 6(6), 222-228

Bell SD, Cairns SS, Robson RL and Jackson SP (1999) Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. *Mol. Cell.* 4(6), 971-982

Bell SD, Magill CP and Jackson SP (2001) Basal and regulated transcription in Archaea. *Biochem. Soc. Trans.* 29(4), 392-395

Birnboim HC and Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7(6), 1513-1523

Bolotin A, Quinquis B, Sorokin A and Ehrlich SD (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(8), 2551-2561

Booyaratanakornkit B, Miao L and Clark D (2007) Transcriptional responses of the deep-sea hyerthermophile *Methanocaldococcus jannaschii* under shiftig extremes of temperature and pressure. *Extremophiles* 11(3), 495-503

Bornberg-Bauer E, Rivals E and Vingron M (1998) Computational approaches to identify leucine zippers. *Nucleic Acids Res.* 26(11), 2740-2746

Botsford JL and Harman JG (1992) Cyclic AMP in prokaryotes. *Microbiol. Rev*. 56(1), 100-122

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254

Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* 10(2), 102-109

Brock TD, Brock KM, Belly RT and Weiss RL (1972) *Sulfolobus*: A new genus of sulphur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84(1), 54-68

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964

Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuño L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Guttierez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Martin C, Mokrousov I, Narvskaïa O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rüsch-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N and Sola C (2006) Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* 6, 23

Brunner NA, Brinkmann H, Siebers B and Hensel R (1998) NAD+-dependent glyceraldehyde-3-phosphate dehydrogenase from *Thermoproteus tenax*: The first identified archaeal member of the aldehyde dehydrogenase superfamily is a glycolytic enzyme with unusual regulatory properties. *J. Biol. Chem.* 273(11), 6149-6156

Brunner NA, Siebers B and Hensel R (2001) Role of two different glyceraldehydes-3-phosphate dehydrogenases in controlling the reversible Embden-Meyerhof-Parnas pathway in *Thermoproteus tenax*: Regulation on protein and transcript level. *Extremophiles* 5(2), 101-109

Burgess RR (1991) Use of polyethylenimine in purification of DNA-binding proteins. *Methods Enzymol*. 208, 3-10

Carstens CP and Waesche A (1999) Codon bias-adjusted BL21 derivatives for protein expression Strategies. *Newsletter 12*

Carte J, Wang R, Li h, Terns RM and Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev*. 22(24), 3489-3496

Chang A, Scheer M, Grote A, Schomburg I and Schomburg D (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res*. 37(Database issue), D588-D592

Charlebois RL, She Q, Sprott DP, Sensen CW and Garrett RA (1998) Sulfolobus genome: from genomics to biology. *Curr. Opin. Microbiol*. 1(5), 584-588

Chomczynski P, Mackey K, Drews R and Wilfinger W (1997) DNAzol: a reagent for the rapid isolation of genomic DNA. *Biotechniques* 22(3), 550-553

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B and Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283-1287

Cole C, Barber JD and Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 36(Web server issue), W197-W201

Combet C, Blanchet C, Geourin C and Deleage G (2000) NPS@: network protein sequence analysis. *Trends. Biochem. Sci.* 25(3), 147-150

Cordin O, Banroques J, Tanner NK and Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* 367, 17-37

Crawford JT (2003) Genotyping in contact investigations: a CDC perspective. *Int. J. Tuberc. Lung Dis.* 7, S453-S457

Cubeddu L and White MF (2005) DNA damage detection by an archaeal single-stranded DNA-binding protein. *J. Mol. Biol*. 353(3), 507-516

Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G and Yang R (2008) Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS One*., 3(7), e2652

DeBoy RT, Monogodin EF, Emerson JB and Nelson KE (2006) Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. *J. Bacteriol*. 188(7), 2364-2374

Demir V and Dincturk HB (2006) Semi-anaerobic growth conditions are favoured by some *Escherichia coli* strains during heterologous expression of some archaeal proteins. *Mol. Biol. Rep*. 33(1), 59-63

DeLong EF and Pace NR (2001) Environmental diversity of Bacteria and Archaea. *Syst. Biol.* 50(4), 470-478

Deveau H, Barrangou R, Garbeau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineua S (2007) Phage response to CRISPR encoded resistance in *Streptococcus thermophilus. J. Bacteriol*. 190(4), 1390-1400

DiRuggiero J, Santangelo N, Nackerdien Z, Ravel J and Robb FT (1997) Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol*. 179(14), 4643-4645

Dodd IB and Egan JB (1990) Helix-turn-helix DNA-binding motifs prediction: Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acid Res*. 18(17), 5019-5026

Enger-Blum G, Meier M, Frank J and Müller GA (1993) Reduction of background problems in non-radioactive northern and Southern blot analyses enables higher sensitivity than 32P-based hybridizations. *Anal. Biochem.* 210(2), 235−244

Enkhbayar P, Kamiya M, Osaki M, Matsumoto T and Matsushima N (2004) Structural principles of leucine-rich repeat (LRR) proteins. *Proteins* 54(3), 394-403

Fadiel A, Lithwick S, Ganii G and Scherer SW (2003) Remarkable sequence signatures in archaeal genomes. *Archaea* 1(3), 185-190

Finn RD, Tate J, Mistry J, Coggill PC, Sammut JS, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL and Bateman A (2008) The Pfam protein families database. *Nucleic Acid Res*. 36(Database issue), D281-D288

Fischer F, Zillig W, Stetter KO and Schreiber G (1983) Chemolithoautotrophic metabolism of anaerobic extremly thermophilic archaebacteria. *Nature* 301, 511-513

Fitz-Gibbon ST, Ladner H, Kim UJ, Stetter KO, Simon MI and Miller JH (2002) Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci. USA* 99(2), 984-989

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31(13), 3784-3788

Georgiou G and Valax P (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7(2), 190-197

Godde JS and Bickerton A (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol*. 62(6), 718-729

Götz D, Paytubi S, Munro S, Lundgren M, Bernander R and White MF (2007) Response of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol*. 8(10), R220

Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. *Gene* 324, 15-34

Greve B, Jensen S, Brügger K, Zillig W and Garrett RA (2004) Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* 1(4), 231-239

Grill S, Gualerzi CO, Londei P and Bläsi U (2000) Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J*. 19(15), 4101-4110

Grissa I, Vergnaud G and Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*., 35 (web server issue), W52-57

Grissa I, Vergnaud G and Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172

Grissa I, Vergnaud G and Pourcel C (2008) CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*., 36 (Web server issue), W145-148

Groenen PM, Bunschoten AE, van Soolingen D and van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol*. 10(5), 1057-1065

Gruber AR, Lorenz R, Bernhart SH, Neuböck R and Hofacker IL (2008) The Vienna RNA Websuite. *Nucleic Acids Res*. 36(Web server issue), W70-W74

Grundy FJ and TM (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit. Rev. Biochem. Mol. Biol.* 41(6), 329-338

Guidotti LG and Chisari FV (2001) Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu. Rev. Immunol*. 19, 65-91

Häring M, Peng X, Brügger K, Rachel R, Stetter KO, Garrett RA and Prangishvili D (2004) Morphology and genome organization of the virus PSV of the hyperthermo-

philic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the Globuloviridae. *Virology* 323(3), 233-242

Haft DH, Selengut J and White O (2003) The TIGRFAMs database of protein families. *Nucleic Acid Res*. 31(1), 371-373

Haft DH, Selengut J, Mongodin EF and Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PloS Comput. Biol*. 1(6): e60

Hale C, Kleppe K, Terns RM and Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14(12), 2572-2579

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM and Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139(5), 945-956

Han D and Krauss G (2009) Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett*. 583(4), 771-776

Han D, Lehmann K and Krauss G (2009) SSO1450 – a CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA. *FEBS Lett*. 583(12), 1928-1932

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166(4), 557-580

Hannon GJ (2002) RNA interference. *Nature* 418, 244-251

Hardy CD and Martin PK (2008) Biochemical characterization of DNA-binding proteins from *Pyrobaculum aerophilum* and *Aeropyrum pernix*. *Extremophiles* 12(2), 235-246

Hartig JV, Tomari Y and Förstemann K (2007) piRNAs - the ancient hunters of genome invaders. *Genes Dev*. 21(14), 1707-1713

Held NL and Whitaker RJ (2009) Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol*. 11(2), 457-466

Ho SN, Hunt HD, Horton RM, Pullen JK and Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1), 51-59

Hochheimer A, Hedderich R and Thauer RK (1999) The DNA binding protein Tfx from *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcriptional regulation. *Mol. Microbiol*. 31(2), 641-650

Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineua S, Boyaval P, Fremaux C and Barrangou R (2008), Diversity, activity and evolution of CRISPR loci in *Streptococcus thermophilus. J. Bacteriol*. 190(4), 1401-1412

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC and Stetter KO (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63-67

Ishino Y, Shinagawa H, Makino K, Amemura M and Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol*. 169(12), 5429-5433

Jansen R, Embden JD, Gaastra W and Schouls, LM (2002) Identification of a novel family of sequence repeats among prokaryotes. *OMICS* 6(1), 23-33

Jansen R, Embden JD, Gaastra W and Schouls, LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43(6), 1565-1575

Jinek M and Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457, 405-412

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen h, Shaw R, Goval M and van Embden J (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol*. 35(4), 907-914

Keeling PJ and Doolittle WF (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci. USA* 92(13), 5761-5764

Kinch LN, Ginalski K, Rychlewski L and Grishin NV (2005) Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acid Res*. 33(11), 3598-3605

Koonin EV and Galperin MY (2003) Sequence, Evolution, Function – Computational approaches in comparative genomics. *Kluwer Academic Publishers*

Kube J, Brokamp C, Machielsen R, van der Oost J and Märkl H (2006) Influence of temperature on the production of an archaeal thermoactive alcohol dehydrogenase from *Pyrococcus furiosus* with recombinant *Escherichia coli*. *Extremophiles* 10(3), 221-227

Kunin V, Sorek R and Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*. 8(4), R61

Kyrpides NC and Ouzounis CA (1999) Transcription in archaea. *Proc. Natl. Acad. Sci. USA* 96(15), 8545-8550

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) ClustalW and ClustalX version 2. *Bioinformatics* 23(21), 2947-2948

Lau NC, Lim LP, Weinstein EG and Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* 294, 858-862

Lee SJ, Engelmann A, Horlacher R, Qu Q, Vierke G, Hebbeln C, Thomm M and Boos W (2003) TrmB, a Sugar-specific Transcriptional Regulator of the Trehalose/Maltose ABC Transporter from the Hyperthermophilic Archaeon *Thermococcus litoralis*. *J. Biol. Chem*. 278(2), 983-990

Lee SJ, Surma M, Seitz S, hausner W, Thomm M and Boos W (2007) Differential signal transduction via TrmB, a sugar sensing transcriptional repressor of *Pyrococcus furiosus*. *Mol. Microbiol*. 64(6), 1499-1505

Lilie H, Schwarz E and Rudolph R (1998) Advances in refolding of proteins produced in *E.coli*. *Curr. Opin. Biotechnol.* 9(5), 497-501

Lillestøl RK, Redder P, Garrett RA and Brügger K (2006) A putative viral defence mechanism in archaeal cells. *Archaea* 2(1), 59-72

Lillestøl RK, Shah SA, Brügger K, Redder P, Phan H, Christiansen J and Garrett RA (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol. Microbiol.* 72(1), 259-272

Londei P (2005) Evolution of translation initiation: new insights from the archaea. *FEMS Microbiol. Rev*. 29(2), 185-200

Mackey K, Williams P, Seim S and Chomczynski P (1996) The use of DNAzol® for the rapid isolation of genomic DNA from whole blood. *Biomed. Products Supplement* 263, 13-15

Mahillon J and Chandler M (1998) Insertion sequences. *Microbiol. Mol. Biol. Rev*., 62(3), 725-774

Makarova KS and Koonin EV (2003) Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol.* 4(8), 115

Makarova KS, Aravind L, Grishin NV, Rogozin IB and Koonin EV (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res*., 30(2), 482-496

Makarova K, Grishin N, Shabalina S, Wolf Y and Koonin E (2006) A putative RNA-interference-based immune system in prokaryotes: computationl analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanism of action. *Biol. Direct* 1, 7

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH (2009) CDD: specific functional annotation with the Conserved Domain Database *Nucleic Acids Res.* 37(Database issue), D205-210

Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Mayromatis K, Ivanova N and Kyrpides NC (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res.,* 34(Database issue), D344-D348

Markowitz VM, Mavromatis K, Ivanova N, Chen IM, Chu K and Kyrpides NC (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25(17), 2271-2278

Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845

Masepohl B, Görlitz K and Böhme H (1996) Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium *Anabena* sp. PCC 7120. *Biochim. Biophys. Acta* 1307(1), 26-30

McClelland M, Chada K, Welsh J and Ralph D (1993) Arbitrary primed PCR fingerprinting of RNA applied to mapping differentially expressed genes. EXS 67, 103-115

McKay DB and Steitz TA (1981) Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. *Nature* 290, 744-749

Mojica FJ, Ferrer C, Juez G and Rodriguez-Valera F (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol*. 17(1), 85-93

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J and Soria E (2005) Intervening sequences of regulatory spaced prokaryitic repeats derive from foreign genetic elements, *J. Mol. Evol.* 60(2), 174-182

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J and Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(3), 733-740

Mullis KB, Faloona S, Sarf R, Saiki RK, Horn G and Ehrlich H (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium of Quantitative Biology* 51, 263-273

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC and Fraser CM (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323-329

Neumann H and Zillig W (1990) Structural variability in the genome of the *Thermoproteus tenax virus* TTV1. *Mol. Gen. Genet*. 222(2-3), 435-437

Ogata H, Audic S, Barbe V, Artiguenave F, Fournier PE, Raoult D and Claverie JM (2000) Selfish DNA in protein-coding genes of Rickettsia. *Science* 290, 347-350

Ogata N and Miura T (2000) Elongation of tandem repetitive DNA by the DNA polymerase of the hyperthermophilic archaeon *Thermococcus litoralis* at a hairpin-coil transitional state: a model of amplification of a primordial simple DNA sequence. *Biochemistry* 14(39), 13993-14001

Ortmann AC, Brumfield SK, Walther J, McInnerney K, Brouns SJ, van de Werken HJ, Bothner B, Douglas T, van der Oost J and Young MJ (2008) Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus turreted icosahedral virus*. *J. Virol*. 82(10), 4874-4883

Pan HQ, Wang YP, Chissoe SL, Bodenteich A, Wang Z, Iyer K, Clifton SW, Crabtree JS and Roe BA (1994) The complete nucleotide sequences of the SacBII Kan domain of the P1 pAD10-SacBII cloning vector and three cosmid cloning vectors: pTCF, svPHEP, and LAWRIST16. *Genet. Anal. Tech. Appl.* 11(5-6), 181-186

Pearson WR and Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* USA 85(8), 2444-2448

Plagens A, Hagemann A, Tjaden B, Neuhaus K and Hensel R (2010a) The CRISPR/Cas system of *Thermoproteus tenax*: Influence of abiotic stress conditions. In preparation.

Plagens A, Tjaden B, Eling H and Hensel R (2010b) Two Cas core complexes identified in *Thermoproteus tenax.* In preparation.

Pourcel C, Salvignol G and Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653-663

Price EP, Smith h, Huygens F and Giffard PM (2007) High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. *Appl. Environ. Microbiol*. 73(10), 3431-3436

Promega Technical Manual (1994) Alter sites® II *in vitro* mutagenesis systems. 316B, 161-169

Riehle MM, Bennett AF and Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 98(2), 525-530

Riley PA (1994) Free radicals in biology: oxidative stress and the effects of ionizing radiation. *Int. J. Radiat. Biol*. 65(1), 27-33

Roche Applied Science Manual (2003) DIG Application Manual for filter hybridization.

Rudolph R and Lilie H (1996) *In vitro* folding of inclusion body proteins. *FASEB J.* 10(1), 49-56

Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, Mullis KB and Ehrlich HA (1988) Primer-directed enymatic amplification of DNA with a thermostable polymerase. *Science*, 239, 487-490

Sambrook J, Fritsche EF and Maniatis T (1989) Molecular Cloning - A Laboratory Manual, 2nd Ed.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* USA 74(12), 5463-5467

Santangelo TJ, Cubonova L, Skinner KM and Reeve JN (2009) Archaeal intrinsic transcription termination in vivo. *J. Bacteriol*. 191(22), 7102-7108

Schleper C, Holz I, Janekovic D, Murphy J and Zillig W (1995) A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J. Bacteriol*. 177(15), 4417-4426

Schneider KL, Pollard KS, Baertsch R, Pohl A and Lowe TM (2005) The UCSC Archaeal Genome Browser. *Nucleic Acids Res.* 34(Database issue), D407-D410

Schramm A, Siebers B, Tjaden B, Brinkmann H and Hensel R (2000) Pyruvate kinase of the hyperthermophilic crenarchaeote *Thermoproteus tenax*: physiological role and phylogenetic aspects. *J. Bacteriol.*, 182(7), 2001-2009

Shah SA, Hansen NR and Garrett RA (2009) Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans*. 37(1), 23-28

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW and van der Oost J (2001) The complete genome sequence of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci.* USA 98(14), 7835-40

Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S and Yokoyama S (2007) Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol*. 189(10), 3891-3901

Siebers B and Hensel R (1993) Glucose catabolism of the hyperthermophilic archaeum *Thermoproteus tenax*. *FEMS Microbiol. Letters*, 111, 1-8

Siebers B, Wendisch VF and Hensel R (1997) Carbohydrate metabolism in *Thermoproteus tenax*: *in vivo* utilization of the non-phosphorylative Entner-Doudoroff pathway and characterization of its first enzyme, glucose dehydrogenase. *Arch. Microbiol.*, 168(2), 120-127

Siebers B, Tjaden B, Michalke K, Dörr C, Ahmed H, Zaparty M, Gordon P, Sensen C, Zibat A, Klenk HP, Schuster SC and Hensel R (2004) Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. *J. Bacteriol.* 186(7), 2179-2194

Siebers B, Raddatz G, Zaparty M, Tjaden B, Plagens A, Albers SV, Bell SD, Blombach F, von Jan M, Kletzin A, Rampp M, Makarova K, Klenk HP, Schuster SC and Hensel R (2010) The complete genome sequence of *Thermoproteus tenax*: a physiologically versatile type strain of Crenarchaeota. In preparation.

Slupska MM, King AG, Fitz-Gibbon S, Besemer J, Borodovsky M and Miller JH (2001) Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*. *J. Mol. Biol*. 309(2), 347-360

Soppa J (1999) Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol. Microbiol.* 31(5), 1589-1592

Sorek R, Kunin V and Hugenholtz P (2008) CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Rev. Microbiol*. 6(3), 181-186

Staynov DZ, Pinder JC and Gratzer WB (1972) Molecular weight determination of nucleic acids by gel electrophoresis in non-aqueous solution. *Nat. New Biol.* 235(56), 108-110

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28(6), 1102-1104

Studier FW and Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189(1), 113-130

Sturino JM and Klaenhammer TR (2006) Engineered bacteriophage-defence systems in bioprocessing. *Nat. Rev. Microbiol*. 4(5), 395-404

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J and Hüttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci.*, USA 99(11), 7536-7541

Tang TH, Polacek N, Zymicki M, Huber H, Brugger K, Garrett R, Bachellerie JP and Hüttenhofer A (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus. Mol. Microbiol*. 55(2), 469-481

Thein SL and Wallace RB (1986) The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. *Human genetic diseases*, a practical approach, IRL Press

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22), 4673-4680

Tjaden B, Plagens A, Dörr C, Siebers B and Hensel R (2006) Phosphoenolpyruvate synthetase and pyruvate phosphate dikinase of *Thermoproteus tenax*: key pieces in the puzzle of archaeal carbohydrate metabolism. *Mol. Microbiol.* 60(2),287-298

Tolstrup N, Sensen CW, Garrett RA and Clausen IG (2000) Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* 4(3), 175-179

Torarinsson E, Klenk HP and Garrett RA (2005) Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* 7(1), 47-54

Treangen TJ, Abraham AL, Touchon M and Rocha EP (2009) Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev*. 33(3), 539-571

Umetsu M, Tsumoto K, Ashish K, Nitta S, Tanaka Y, Adschiri T and Kumagai I (2004) Structural characteristics and refolding of *in vivo* aggregated hyperthermophilic Archaeon proteins. *FEBS Lett.* 557(1-3), 49-56

Valentine DL (2007) Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature Rev. Microbiol.* 5(4), 316-323

van der Oost J, Jore MM, Westra ER, Lundgren M and Brouns SJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci*. 34(8), 401-407

Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T and Itakura K (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res.* 6(11), 3543-3557

Waters LS and Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136(4), 615-628

Weber K and Osborn M (1969) The reliability of molecular weight determinations by dodecyl sulphate-polyacrylamide gel electrophoresis. *J. Biol. Chem*. 244(16), 4406-4412

Werner F and Weinzierl RO (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell*. 10(3), 635-646

Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W and Doudna JA (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17(6), 904-912

Williams E, Lome T, Savas J and DiRuggiero J (2006) Microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus* exposed to gamma irradiation. *Extremophiles* 11(1), 19-29

Woese CR and Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci.* USA 74(11), 5088-5090

Woese CR, Kandler O and Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* USA 87(12), 4576-4579

Yakunin AF, Proudfoot M, Kuznetsova E, Savchenko A, Brown G, Arrowsmith CH, Edwards AM (2004) The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2',3'-cyclic phosphodiesterase, 2'-nucleotidase, and phosphatase activities. *J. Biol. Chem*. 279(35), 36819-36827

Zegans ME, Wagner JC, Cady KC, Murphy DM, Hammond JH and O'Toole GA (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviours of *Pseudomonas aeruginosa*. *J. Bacteriol*. 191(1), 210-219

Zillig W, Stetter KO and Janekovic D (1979) DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. *Eur. J. Biochem*. 96(3), 597-604

Zillig W, Stetter KO, Schäfer W, Janekovic D, Wunderl S, Holz I and Palm P (1981) *Thermoproteales*: a novel type of extremely thermoacidophilic anaerobic archaebacteria isolated from Icelandic solfatares. *Zentbl. Bakteriol. Hyg.*, 1 Abt Org C 2, 205-227

Zillig W, Prangishvilli D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D and Götz D (1996) Viruses, plasmids and other genetic elements of the thermophilic and hyperthermophilic Archaea. *FEMS Microbiol. Rev.* 18(2-3), 225-236

LIST OF ABBREVIATIONS

A. bidest.	aqua bidestillata = two times distilled water
aa	amino acid
Amp^r	ampicillin resistence
APS	ammonium persulfate
bp	base pair(s)
α-Me	beta-mercaptoethanol
BSA	bovine serum albumin
Cam^r	chloramphenicol resistence
CasA1	Cas complex in Archaea 1 (Cas4, Cas1/2, Csa1)
CasA2	Cas complex in Archaea 2 (Csa5, Csa2, Cas5a, Cas3, Cas3HD, Csa4)
Cas	CRISPR-associated protein
CDP	disodium 2-chloro-5 (4-methoxyspiro{1.2-dioxetane-3.2'-(5'-chloro) tricycle [3.3.1.1.$^{3.7}$] decan}-4-yl)-1-phenylphosphate
CE	crude extract
CIAP	calf intestinal alkaline phosphatase
Conc.	concentration
CRISPR	Clustered regularly interspaced short palindromic repeats
Csa	CRISPR-associated protein, subtype *Aeropyrum pernix*
Csm	CRISPR-associated protein, subtype *Mycobacterium tuberculosis*
DEPC	diethylpyrocarbonate
DIG-UTP	digoxigenin-labeled UTP
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dsDNA	double-stranded DNA
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen = German Collection of Microorganisms and Cell Cultures
DTT	dithiothreitol
e.g.	for example
EDTA	ethylene-diamine-tetraacetic acid

EMP	Embden-Meyerhof-Parnas pathway
EMSA	Electrophoretic Mobility Shift Assay
et al.	et alteri = and others
Fig.	figure
g	gram
x *g*	gravitational acceleration
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
(6x) His-tag	(hexa) histidine tag
HGT	horizontal gene transfer
HP	heat precipitation
HTH	helix-turn helix motif
i.e.	id est = that is, that is to say
IPTG	isopropyl-ß-D-thiogalactopyranoside
Kan^r	kanamycin resistence
kb	kilobases
kDa	kilodalton
l	liter
LB	Luria-Bertani
M	molar (mol/l)
m	milli (10^{-3})
µ	micro (10^{-6})
mA	milliampere
min	minute
MOPS	3-(N-morpholino)propanesulphonic acid
MW	molecular weight
n	nano (10^{-9})
NAD+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
NCBI	National Center for Biotechnology Information
Ni-NTA	nickel-nitrilotriacetic acid
OD	optical density
ORF, *orf*	open reading frame
p.a.	pro analysi

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEI	Polyethylenimine
pH	negative logarithm of the hydrogen ion (H^+) concentration
Psi	pound-force per square inch
Pfu-Polymerase	DNA-polymerase from *Pyrococcus furiosus*
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
rNTP	ribonucleoside triphosphate
rpm	rounds per minute
RT	room temperature
RT-PCR	Reverse-Transcription PCR
s.	see
SDS	sodiumdodecylsulfate
sec	seconds
sp.	species
SSC	standard saline citrate
ssDNA	single-stranded DNA
Tab.	table
TAE	tris-acetate-EDTA buffer
Taq-Polymerase	DNA-Polymerase from *Thermus aquaticus*
TBE	tris-Borate-EDTA buffer
TBP	TATA binding protein
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFB	transcription initiation factor B
T_m	melting temperature
Tris	tris-(hydroxymethyl)-aminomethane
U	(enzyme) unit
UV	ultraviolet
V	volt
Vol	volume
W	watt
www	world wide web

x	fold
<	lower than
>	higher than
% (v/v)	percent by volume
% (w/v)	percent by weight

Amino acids (1-letter code)

A	alanine	I	isoleucine	R	arginine
C	cysteine	K	lysine	S	serine
D	aspartic acid	L	leucine	T	threonine
E	glutamic acid	M	methionine	V	valine
F	phenylalanine	N	asparagine	W	tryptophan
G	glycine	P	proline	Y	tyrosine
H	histidine	Q	glutamine		

Nucleosides and nucleotides

A	adenosine	AM/D/TP	adenosine mono/di/triphosphate
C	cytidine	CTP	cytidine triphosphate
G	guanosine	GTP	guanosine triphosphate
T	thymidine	TTP	thymidine triphosphate
U	uridine	UD/TP	uridine di/triphosphate

IUPAC-Code of nucleic acids

B	G/T/C	M	A/C	V	G/A/C
D	G/A/T	N	A/C/G/T	W	A/T
H	A/T/C	R	A/G	Y	T/C
K	G/T	S	G/C		

APPENDIX

A1: Nucleotide sequences of CRISPR loci in *T. tenax*

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

A3: Detailed identification of *T. tenax cas* genes

A4: Nucleotide and corresponding amino acid sequences of the *T. tenax* operons *casa1* (TTX_1245-1248) and *casa2* (TTX_1250-1255) and the Csa3 protein (TTX_1249)

A1: Nucleotide sequences of CRISPR loci in *T. tenax*. CRISPR loci identified via CRISPRfinder, leader sequences (blue), repeats (red), spacers (black) and degenerated repeat (green) marked. Comparison of the AT-content of both flanking sites for identification of leader sequence site (DNA Stats).

TTX_1 Thermoproteus tenax: 227260-224864

GCACTACATCCTCTCAACACTGAGTTTCAAGACCAGAAACCTAGATAACTGAAAGGAGATCTATATAGACCACGG
CGTTTATAAATCGTCGTGCAACTTGCCCTTTCCTTCTGTAAAATCGTTCTACTCCTCCGCCGTTGTGTTTTTGTC
AAGCGGGGATCCCGCATATAAATCGTCGTGCGGAGGCGCGGCGGATGGCGGCGCCGAACCCCAAAGACACACTTT
CAACTATCTGAGACAGAAGCCGGCCGCGGAAAATAGACAAAACTGCTATCATGCCACCTCAAGAGTCCGCCGAGA
TCAACGTCAGTCTCGTGAGAAAAAAGCTTAAAAGCTCGTAGAAACAAAGACAACAATACCCGGAATCTCAGATAG
AGATTTGAAGGCCGGCGCCGACATTGATGCATATGTTGTTCACATCGAGATAGTGAATCTCAGATAGAGATTTGA
AGGTTTCTCTTTCAATTTTTCTCCCTAATCTTCTCCCTAAACAGAATCTCAGATAGAGATTTGAAGGCGGCGCAC
AGTCTGTATCGTTCACCCCTGTGAGCGGCTATACGAATCTCAGATAGAGATTTGAAGGTCGGTGATGGAGTACCT
GTCTATGCCCGCCTCATCGAGCACTCTGGGAATCTCAGATAGAGATTTGAAGGCTTAGTATAGGCCAAATGACAA
CTAGCAGGAACACTATCAGCAATGAATCTCAGATAGAGATTTGAAGGAACTTCTTCGTCGTGGAGGGGCCGGACG
GCCGGCAATACAGGGAATCTCAGATAGAGATTTGAAGGTTGCCGGCGTCTGTCAGCACGGCCGTGGTTTCCCTCC
AATTGAATCTCAGATAGAGATTTGAAGGTTGGACGCTGTGTTGGCATGCGCCGACCCGCGGCGGATTGCTGAATC
TCAGATAGAGATTTGAAGGATTGGATATATGTGCGGCATCCTCGTATTGCGAAATACGCTCCAGAATCTCAGATA
GAGATTTGAAGGGAAAGCCCAACCTAAGCTTCGTCTCAAGCATGGCGAGATCGTGAATCTCAGATAGAGATTTGA
AGGTACCAGGTGCAGTGCAACGGCATCAGCTATACCGACTACACCTGAATCTCAGATAGAGATTTGAAGGCAACT
GAGCATCACGCAGGGGGCGGTCGTCGACGGAGGGTACCTCTACTCGAATCTCAGATAGAGATTTGAAGGACCCAG
TAGGATATCGCCTCATAGCGGTATTTGAAATCGCCGAATCTCAGATAGAGATTTGAAGGTACGAATATATCGGAC
GCAAGGTCGTAAAGTTCGAGCTCTTCTGGAATCTCAGATAGAGATTTGAAGGTCTTTTGTCACTATCACAGCCAA
TCCCGAATCGAGACGGACCGAATCTCAGATAGAGATTTGAAGGCTTATCCTGACCGGCTTCGCGTACGGCTTGTT
CACATATCTGCACGAATCTCAGATAGAGATTTGAAGGTTTCTGAAGGAACGGGTGAAGGAGGTGCTGAAGAAGTA
CGGCGGAATCTCAGATAGAGATTTGAAGGGCGTAGCTGTTCCTCCGTTATCTTTATTGCGATGTTCTCGAATCTC
AGATAGAGATTTGAAGGCCCTTGCTTAGCGAGCTCATTGGCGACAAGGTTGATCTTGTTATGAATCTCAGATAGA
GATTTGAAGGATATTCGACGAGACGGGCAACACGAAGATCGGCACATGGATCTGAATCTCAGATAGAGATTTGAA
GGCAGATGCGCCAAGCTATCGAGTCCACGTGTCCAACCGAGCGGAATCTCAGATAGAGATTTGAAGGTACGTGCG
CGTTGAGCGCGATATGAAGTTCGACGCAGTCGAGGAATCTCAGATAGAGATTTGAAGGAGGAGGTCGCCGTCCCA
AGATGTATTGCTTCGACTATGAGCAAGGAATCTCAGATAGAGATTTGAAGGGGGTTCCAGCCAAGCGCATGCGCC
AATCCCCATCCAGCTATTGAATCTCAGATAGAGATTTGAAGGGGGATATAGATTTGGAAAGATCAAAATATTTAA
GAACATTTGTTCTACATTTTTTATGAATAACAAAACTGAACAAAAGAAACTATCCTTGGAACCGTTGGTTCACGT
CTTGTCGTGTCTGGCGCGTGAGAAGTACTATGGCCCGCTTGACCGCCTAGCCCGCGTTGAGGACAACGACGAGTT
CTACGTCACGACAGCTATGGAGGCCCTCTACGATGCGTTGAGGTACTTCGGCACACTGGAGGGCTGTGCCCGTCC
GGAGGGGCTGGATGAGGCTGTCCGCGCCTTCGCCGCGTTGGTTAAGAGGAGACCGGCTGTGGCGAAGGAGGTGGC
GTTTAAGGCGCTTGCGAAGGCGGTGGGTGGTGAGCGGCAGGCGGTGAAGGGACGGGCGTCTGCGGCGAGGTAG

	Flanking site 1 (362 bp)	Flanking site 2 (406 bp)
g,c	176, 48.62 %	224, 55.17 %
a,t	186, 51.38 %	182, 44.83 %

TTX_2 Thermoproteus tenax: 316783-317791

AACGCGGCCGAAGTAGAGGTACACATCGATCACAACAAAGGGCGAATCGCGCCGCTACATATATGTCAGTAAAGG
ACATAATCGACATTTAAAGAACGCGTTGGAAAGGCAAGGCGAAGAACGGCGCTGGGTCGCGCCTTCTCTAGGGCG
CAGTCGTAAAGGCGAAGAACGGCGCATGCCGTAGCCCGGTCAGAGGAATATATGAATGTAAATGGGCGTTTGAAG
GCGACCGCCGGGCGGGCGGAGGATCTGGGCAAAATTTAAATAGCCGGGGCCGCAATCGACGTGGGCGGTGGAAAT
CAAAAGATAGTAGAAACTGTACTTTAGGAATAATATACTCGTCAAGTATCTTGTAATACAGTGGAAATCAAAAGA
TAGTAGAAACTTGCCCTTCAGAAGGTCGCCCTCCTTTACGTTGCAACCCCAAGTGGAAATCAAAAGATAGTAGAA
ACATTAGACGTTGTAGCACTAACGCCTCTTTCACAACATTTATAGTGGAAATCAAAAGATAGTAGAAACGCCGGG
CATGTGAGCCCCCGTTCGATGTGCTCATAAGTCTGCGTGGAAATCAAAAGATAGTAGAAACGATCAAAAAAGGAG
GTGGACGTTCAACGTCCTCTGCACGTACCTAGGTGGAAATCAAAAGATATGTAGAAACATACACATCGGCACCCC
ATGCTATGTATTTTTCTCGCGTGGTAGTGGAAATCAAAATATAGTAGAAACAGATTTCCGCATCAACACAAAATA
GAAACAGAAACCCCTATATTGCCGGCTAAGCGCGGCGCCGGATATATTTCAGGAGCGTTCTCCCCGCCTCCGTGA
GTCCGGCGTCGTTTAAAAGTCCGTGTTCTGCAAATTCCCTGAGAATCCTTCTGTCGTACCGCCCGAGGTCCTCTA
GAGGGCCTTTGGCCAGCTCCCTCCTCCTGGCGCTCGCGTTGGAGCGCGTGAGCTCGTTGACGTAGAGTATCGGCA
TTATCTGGGCTATCCTCTCGGCGTGCTCCGCCAT

	Flanking site 1 (292 bp)	Flanking site 2 (284 bp)
g,c	162, 55.48 %	161, 56.69 %
a,t	130, 44.352 %	123, 43.31 %

TTX_3 Thermoproteus tenax: 346121-345073

GATATCCCACGAGCATTAAAACATTCCAATATATGTCGGTAAAGGATGTAATATACTCCAGACAAGGGGCTAGAT
GGCCAAATGCGAAGAACGGCGCTGGGTAGCCGCCGTCTCTAAGGCATTGTCGCGATCGTAAAGGCGAAGAACGGC
GCACGCCGTAGCCCGGACAGGAGAATGTGTGAATGTAAAGGGCAGCGGAAAAGCGGCCGCCGGGCGGGCGGAGGA
TCTGGGCAAAATTTAAATACCCGAGGCCGCAGTCAACGTGGGCGGTGGAAATCAAAAGATAGTAGAAAGCTTCCA
CAAAGTGAGCACCGCCAGTCTTAGAAATGTAGCCTAGCTGTGGAAATCAAAAGATAGTAGAAAGGTTATCTCCGA
TGGACTGGCCACACACAGAGAGAGGTAGGTGCGTGGAAATCAAAAGATAGTAGAAAGATCACGTTTGACTTCCAC
AGTTACGTGGTCGGAAGAAAAAGCGGGGGTGGAAATCAAAAGATAGTAGAAAGTACGGGAGGGGATATTTTAGCT
CCCCAGCCGCTCCAGAGACCTCTTCGGTGGAAATCAAAAGATAGTAGAAATGTGAGCCATGGCACTACAGGCCCC
CACGCTGTATCGCCATTGAGTGGAAATCAAAAGATAGTAGAAATATCGCGATGCTATATAGCCTCGCGTGGACGG
CGTAGAGCGCGTGTGGAAATCAAAAGATAGTAGAAACGTTCTAATTCTCGTCTTGATCCTCTCCTCTGCTCCTCA
GGGATGGAAATCGAAAGATAGTAAACGGCACATACGGCTACACCTGCGTTCCTTAGTAGGGTCGCTAAAAACAGC
GCAACGCATGTCATAATAATTGAACGTAGAGACAGCCCTCGTCCAGCGTTGTAGCCGTTATCCTCTGCCGTGGAG
GCGGCTACATTGCATTTGCGTCCACCCGGGACGGCGGACTGATGTGCGCCTCTCCTGGGCGTCGCCAGAGGCCGC
AGACCTGAGCCCCAGCGGGCCGCGCCGGTTCAGCCGCGTGCCCCATCCAAACGACGGCGTGCCCCTCTGCGTCT

	Flanking site 1 (269 bp)	Flanking site 2 (274 bp)
g,c	148, 55.02 %	172, 62.77 %
a,t	121, 44.98 %	102, 37.23 %

TTX_4 Thermoproteus tenax: 1078562-1075638

GGGCAGACCGGCGGAGCTCAATTTATGCGTTTCGGGCCCGACCTATATATCCTAAATAATTCTCAGCGTTTAAAA
ATCGTCGTGCAGCTCGCCCTTTCCTTCTGTAAAATCGTTCAACTGCCGCGCCGTTGTATCTCTGTAAAGCGGGGA
TCCCGTTTAAAAAACGGCGGACAGAAGTACGGCGTTCGGCGGCGCCGACCTCCAGCCGCACGCTTTCAACTGCCG
GAGGCGGGATCCGACCGCGTAGAACAGACAAAGCCCTCCCCGGGGCAACACAAGGTCGCGCCGGGACTAACACAA
AGACCCACAGAGGAAAACTTAAAAATCGGCAAAAACTAAAGACCAGAAGGCCCGGAATCTCAAAGAGAGGATTGA
AAGTGTATTGAGAGATGTAGTTGATTATTCGTGCGGCGCAAATGTCCGCGAATCTCAAAGAGAGGATTGAAAGCT
CGCCGCTCACGGGGAGGAGGAGGAAGAGGCCCACGTAGGCGTCGTATAGAATCTCAAAGAGAGGATTGAAAGAAG
TACATCCGCGATCGTTAACCTCTTTTTTCTCATCTTGAATCTCAAAGAGAGGATTGAAAGTGACACAAAGGGCGT
TGACGTCACCGACATCCCCGACGCGAATCTCAAAGAGAGGATTGAAAGTCTTCGCCCCGGCCGTCTGGGGAGCCG
ACTGGCTCCTTCCCAGAATCTCAAAGAGAGGATTGAAAGCGTATTTAACATCACAGCTGTGTTTCAGAGCGCCTC
AGAATCTCAAAGAGAGGATTGAAAGCGAAGTCGGATATGCGATCGCATGCCTGACCACGTAGATCCGCGTGAATC
TCAAAGAGAGGATTGAAAGCTATATACACGTATATACACAGAGGGGCCCTTTGAAGGCGGGAATCTCAAAGAGAG
GATTGAAAGTCGCGGGGGAGCGGATACAAGACATCCTGGCGTGGTACGTCGCGAATCTCAAAGAGAGGATTGAAA
GTGCTTAACCACTCGCCCGGTATTTCCTCGATCTTTTCTTCGAATCTCAAAGAGAGGATTGAAAGTCGAGTCGCT
CTGCGGCGGTGAGCCGCTGTCCGAATGTGTGAATCTCAAAGAGAGGATTGAAAGCCACTGCGCTATATGTTCCAG
GACGTGCTGGAGGATGTGAATCTCAAAGAGAGGATTGAAAGGCTCGTGTACACGCAGCGCAAGCACTCACAGTGT
CTCGTCGGAATCTCAAAGAGAGGATTGAAAGTAGAGTACCACCCAGACGGCCGCATGCTGAGCAAGTGCGACAGA
ATCTCAAAGAGAGGATTGAAAGCCTTGCGGGGCACGCTCCTCAACCCCTCCGTCAGCCCCCTCCAGTCGATGAAT
CTCAAAGAGAGGATTGAAAGACTGGCGGATACACGTTATGCCGATTGGAGAGTCAATATACTTGAATCTCAAAGA
GAGGATTGAAAGGCAGTTCGCCTTGTGGGCCTACTTGATAAACGTGTTCTTCACAGGAATCTCAAAGAGAGGATT
GAAAGTCTTGACTAGATAGCCCTTCTTGCGCTTTTTCTCCGATGGCTGAATCTCAAAGAGAGGATTGAAAGAATC
CCTGTACAGCTCCTCAATATAGCGCACATATTCGCGCCAGAATCTCAAAGAGAGGATTGAAAGTTGAGAATATTC
TATAAAGAATTGCTGAATTTTCTGATGGGCGAATCTCAAAGAGAGGATTGAAAGCGATTTGAGGAAAAAACCTTT
AGGCGCGGCGGCTATGTACTATGAATCTCAAAGAGAGGATTGAAAGTCTTCTCGAGGTATCCCAATCTGCCAGCA
AAGAGCTCGAATCTGAATCTCAAAGAGAGGATTGAAAGCTTTTCCCTCATTCTGTCGACGAGGATCTCCTGGCTT
TCTTCCTGGGAATCTCAAAGAGAGGATTGAAAGAGTCGTGTCCGCTACCACCTCCTAATCCCTTCTTTTTTCGGA
ATCTCAAAGAGAGGATTGAAAGGCGACGATCAACATCGGCACAATTAACATCGATACCATCGTGAAGGAATCTCA
AAGAGAGGATTGAAAGATCTATTGATGACGAAGCTAACAATATCAGCTTGAACACACCATTGGAATCTCAAAGAG
AGGATTGAAAGGGCCTGACGGACAGCACTGAGCCGTCAAAGAGGACGGACAACGAATCTCAAAGAGAGGATTGAA
AGCATGTCACAGGATCATGCTGATAAGAGAAAACATGGAATATACGGAATCTCAAAGAGAGGATTGAAAGGTAAG
TGACAATCTTACAGAGTACTTCTCTATTGCTCTCCATAGAGCTTAAGCACTCGAATCTCAAAGAGAGGATTGAAA
GTAAGACTTGGGCCACAAGTCTTAAGAATGAGGAATAGAAGAATCTCAAAGAGAGGATTGAAAGTTTTCTCTACA
TAGCTTTTCTTTATCGTTATCTCAACACTGAACTGAATCTCAAAGAGAGGATTGAAAGCAACTTCTTTTTCTGTA
TACTCTAATTTGTATGCAAGCGATGAATCTCAAAGAGAGGATTGAAAGTACCAGGAGTTCTGGCTACTCTACTTC
CAGGCGCCGTTTTGGGAATCTCAAAAAGAGGATTGAAAGAGGGCGCTTGATTGTTGGTCTCCTCCTTGGTTGCGA

ATCTCAGGGCGTATTGAAGGCGCATTGATGTTGCTTATGGTGGGGCCTAGTAAGGCGTTGGGGTTCGGCGTGGCG
CTGTGTGGTGTGGGCGTTCGGCCTGCGCTGGGGCTATGGGCACTGGGTTAGTGCGAAGCAGAGGCATCTGTCGTC
GATTGTGGCGATGCAGCCCCCCGTGGTTGCCCCTTCTCTGTAGTCTATTTGGCATTCCGCGTGTCGGTAGGTGCA
GTCGTGTTTCGCAATTAGGAAGAGGTCGTCCCGTTGGCACTGGCAGAGGGCGGCGTAGAGCTCGTGGTATTGGTA

	Flanking site 1 (354 bp)	Flanking site 2 (336 bp)
g,c	194, 54.80 %	198, 58.93 %
a,t	160, 45.20 %	138, 41.07 %

TTX_5 Thermoproteus tenax: 1081524-1084202
GCTACCACCAACGGACTGTATTAAAGATATACTTCACCGACCTATATATCCTAAATAATTCTCGGCGTTTAAAAA
TCGTCGTGCAACTTGCCCTTTCCTTCTGTAAAATAGTTCTACTCCTCCGCCGTTGTATCTCTGTAAAGCGGGGAT
CCCGCATATAAATCGCCGTGCGGAAGCACGGCGCTCGGCGATGCCGAGGGTCGGACGCACGCTTTCAACTGCCGG
ACGCAGGACCCAACCGTGTAGAACAGGCAAAGCCCTCCCCGGGGCAACACAAGGTCGCGCCGGGACCAAGACAAA
GACCCGCAGAGGAAAACTTAAAAATCGGCAGAAACTAAAGACCAGGAGGCCCGGAATCTCAAAGAGAGGATTGAA
AGTCTGGACTGTGTGTTGCAACAGCGAAGTCGCGAACACCTCGACGAATCTCAAAGAGAGGATTGAAAGCGTTGA
TGTGGCCGGGACTGGCTGACTCAGCTATTACGTTGAATCTCAAAGAGAGGATTGAAAGACTTCAAGGCGTGCTTG
CTGGCGTTTTAGGCGACGATGAGCTGAATCTCAAAGAGAGGATTGAAAGCTCTTTTTCCTGCCCCGTTCTTCCAA
CTCTTCGCCGGTCGTGAATCTCAAAGAGAGGATTGAAAGCCGAGTAGCCGACGGCGCGTCCGCTCACTTAAGATC
TCAGAATCTCAAAGAGAGGATTGAAAGTGTGATAGATGAGATAGACAAAATCAAGTTCAGCACCACGTGAATCTC
AAAGAGAGGATTGAAAGGCGCACTAATTACCGCCCCGGCGCAGGATTACCAATCTTGCCGGGCCGGTTGAATCTC
AAAGAGAGGATTGAAAGTTGGTCTCAGGCGCGCGGTAGCCTTCCAGCTGTACGACCAACTCGATGAATCTCAAAG
AGAGGATTGAAAGTCCGGAGGAGGCGGCGAAGAGGATGATAGAGTTGGCCAGGCTGGGAATCTCAAAGAGAGGAT
TGAAAGGGCATGATTGGTTGCGCCGGGAGCTCTGGGATGTGATCCGCCAGAATCTCAAAGAGAGGATTGAAAGCT
ACTTGGAGTATGCCGGTCAAATTCGGTGGGGCGGACGTCGTACCGAATCTCAAAGAGAGGATTGAAAGTTTGTTG
AAATAATCACCGTAACCCAGGCTTTCCAACCATCTTGAATCTCAAAGAGAGGATTGAAAGATCGCTTGTCCAACC
GGGCTCCTCTATAATTGTCGTCATTAGCTTAGAATCTCAAAGAGAGGATTGAAAGACTGGATCATCGATGGGCAG
AATGCGGGCGTGTCGGCTATATGGAATCTCAAAGAGAGGATTGAAAGCAAGTCGTCGAGAAGTATTTTGGTCTGC
TGAAGCAGTACGGTTTCGGAATCTCAAAGAGAGGATTGAAAGGTACCACGACATCAGCCATGTGACTGTTGCGCC
TAGCTCAATCTCGAATCTCAAAGAGAGGATTGAAAGACTTGAATTTCAACAATGTCACCAGCACATATGTCCAGG
CTAGGGAATCTCAAAGAGAGGATTGAAAGATTTTTATTCCTCATAAACATGGTCTCAAGGTCAGAACTAAGCCGA
ATCTCAAAGAGAGGATTGAAAGAATGGCTCCAAGGGATCGTTCTCTAGACTCCAATTTTTCTAATGAATCTCAAA
GAGAGGATTGAAAGTTTGTCGACGGGATCTTGCGACATGCGGTACGCTGCAGTGTCGGAATCTCAAAGAGAGGAT
TGAAAGGCTCCACCGACCAGTACCCATATAGCGTGAACTCTTTCGCCGGAGAATCTCAAAGAGAGGATTGAAAGA
GTGCGCAGACCCCTGCACAGGCCGGCGTCCTTCACAAAAGAATCTCAAAGAGAGGATTGAAAGATGTAGAGCTTG
ACCGGCGCCCCAGGATAGGGCACCGGGCCGGAATCTCAAAGAGAGGATTGAAAGTCTGTATCCCGCCGGCGCGTG
GAGAAGGCGTTTCTCCACAACGTCGTCCTCGAATCTCAAAGAGAGGATTGAAAGTCGAGGACGATGTAGTCGGGA
TTGCTCGCGTTTATCAGCTTGAGAATCTCAAAGAGAGGATTGAAAGCTCGTGGATACTCCAAGCGAAAGAGTTCA
CGTTATACGGCTACGAATCTCAAAGAGAGGATTGAAAGTCCACAAGCCTCACGGCCGTCTTCACGGCGCGGCCCG
GTCTGACGGGAATCTCAAAGAGAGGATTGAAAGGGGATATACTACACGGCCCTGCGCGTCATCGGGAACAAGCGC
ATGGGAATCTCAAAGAGAGGATTGAAAGTATTCCGTAGATTCCGGCGTAAAAGGTCTGGGCGGCGATGAATCTCA
AAGAGAGGATTGAAAGCCGTGCACCTATGACTTTCAAGTACTCAAGGACGTTAGGAATCTCAAAGAGAGGATTGA
AAGAACGCTTGGACAAATACCGTACCGCTACCCTCATATTGATAGAATCTCAAAGAGAGGATTGAAAGGGGATAT
ACGTCTCGGATTAAGCATAGCAGCTTCACCTCTGCGGAATCTATATGGTTTGAAAGGTCTCGAGTATAGACTACA
CGCCGAACTTTACATACTGAGCCTCAAAGCGAGCTCGTTGAAAGCCGGCGTGCGGCGCTTGTTCCAACGTGAGTT
GGTCTAATGGAGGTCGGGGGCGTTGAAGGGAGTTTAGGCCGGCGCCGTGAGGGG

	Flanking site 1 (353 bp)	Flanking site 2 (211 bp)
g,c	187, 52.97 %	114, 54.03 %
a,t	166, 47.03 %	97, 45.97 %

TTX_6 Thermoproteus tenax: 1097343-1094882
ACGTCCGGTCAGCGGGTCCACGAAATTGGAGAACCTATATATCCTAAATAATTCCCGGCGTTTAAAAATCGTCGT
GCAACTTGCCCTTTCTTTCTGTAAAATCGTTCTACTCCTCCGCCGTTGTGTTTTTGTCAAGCGGGGATCCCGCAT
ATAAATCGCCGTGCAGAACAACGGCGGACGGCGTCGCCGACAGTCAAACACGCGCCTTCAACTGCGGGACGCAGG
GGGCCAGTTCTAGAGAAATCAGCAGTCCGGGGAGAGCTATCTCAAGAACACGCTGCAATCAACGCAAAACCCACC
AGAGAAAAACTTAAAAACAGCCAAAAACCAAAACCCAGAAGGCCCGGAATCTCAAAAAGAGGATTGAAAGATAGC
TATCTAAGCGCAATAGTAGAAATTCCAATTCTAAGCTTTGAATCTCAAAAAGAGGATTGAAAGAGCTCTGGGTAT
TTGACAATTACTTCGGGCTTGGGGATGCTCGTGGAATCTCAAAAAGAGGATTGAAAGCCTCGTTCCTTTTCGCTA
TCACTACATCTCCTAATTTGTTTGAATCTCAAAAAGAGGATTGAAAGCCGCAAGGCCAGTGCATGGATTCTCCGT
ACCTCTCAGTGAACCGAATCTCAAAAAGAGGATTGAAAGTCAGAGCAAGGGGTACAAGGTCTACCAGTACGGGGA

CGGTCTCATAACAGTGAATCTCAAAAAGAGGATTGAAAGTCGCAGTTCTCTATTGTTATTAAGGCCTTATAATTT
GTATATGTTCCCTGAATCTCAAAAAGAGGATTGAAAGGGGAATACATTAAGACCAATAAAAATTGTTCACCTCGG
CTTTGAATCTCAAAAAGAGGATTGAAAGTTGAACGTCACGCCGGCGCTGTCTCAACTACTTCGCGGTGGCTGAAT
CTCAAAAAGAGGATTGAAAGTTCATCGTGGCGGCCTTCCTATTCTGGTGGTGGCTGAAGAAGAATCTCAAAAAGA
GGATTGAAAGGGGACAGGTTCAAGGGGTTCCTCGGCCTCCTCATCGTGCTGAATCTCAAAAAGAGGATTGAAAGG
TTGACCAGGTACTTCGACAACCTTTACACAGCCAGGGAATCTCAAAAAGAGGATTGAAAGGCCGGCCTTCGTCGC
ATGCTCCGACGAGTGGAGCCACGCTGAATCTCAAAAAGAGGATTGAAAGACCTGCATCTCCGACGTTGGGGATGA
CGCGGGCGTCGGAATCTCAAAAAGAGGATTGAAAGTAATGGATCAATATAGAACTCAGTCTCGTGGTTGGTCTCA
TATAGAATCTCAAAAAGAGGATTGAAAGCAGGGCATAAAGCCTATTGCGTTGAAGTGCTTAACTACCTCTTGAAT
CTCAAAAAGAGGATTGAAAGATCTCCGCCTTTATCGCCAGAATTCTGCTGTCCATTTGGGTTTCGGGGAATCTCA
AAAAGAGGATTGAAAGATGAGATCCCACCAGTTGAACGTGCCCGCCTGGCTCGGAATCTCAAAAAGAGGATTGAA
AGTATCGACAAAGCATCTAATGCCACGCTCAACGATATGTATCTCGAATCTCAAAAAGAGGATTGAAAGATCAGC
CGGCTCAAGTCGTTTGGTACGACAACGTAACGATAGAATCTCAAAAAGAGGATTGAAAGTGAACTTTCTACAAAT
ACCTGCATTATATTATCCGTAGAACCTAGAATCTCAAAAAGAGGATTGAAAGCGACAGCCTGACGCGATGATGGC
CGCGGGGAGCCCGGCGGAATCTCAAAAAGAGGATTGAAAGGCGGAGAACGAGTACGTGATAGTTGTCCCGGCCGG
AAGTAGCACGGAATCTCAAAAAGAGGATTGAAAGATTTATTTTTATTCTTTTCTGGATCTGAGTTTGGGCCTTCA
GAATCTCAAAAAGAGGATTGAAAGACCCAAGTACAAAGTCAATGAGGCCGGGGAAATTGAGCTCGGAATCTCAAA
AAGAGGATTGAAAGCGATGGGCGACCTCGAGACGAACTCCTCAAGGACGCGCCTGACCGGCGAATCTCAAGTTGA
GGATTGAAAGCTTCAATCTCATCTAGTGCGTCCTCAAGCCGCATCAGCACATATAGAGTCTTAAATAGAAGATTG
AAAGAGCATGTATGGAACTATGTCCTCGCTTGGCATGTATTCGGCGTTGAATCTCGTGATGGCGGTTGAGGGGGC
AGGTCGCTGTGGAGGAGGGGTTGGTGCCGCGTCGGCGGTGGGGAGTGCGTTGGGATGTATAATAACTTCTGTGTT
GAGGGGGGTATGAGGGTTGTTGTTGAGAATTTTGGCGTTATTAGGGAGCTCAATGTGGAGCTTCCTGAGGTTTTT
CTGCTCTTTGGGCCGAATGGGGTGGGGAAGACCACGATATATACCGCGATTACGGCGTTTGGCTACTTGGAGTGG
GGGAGGGGGCTGTACCTGGGGAGACGTGTCAGCGCTTGGAGGAGGACGCCTCGGGGTCGTG

	Flanking site 1 (346 bp)	Flanking site 2 (357 bp)
g,c	177, 51.16 %	202, 56.58 %
a,t	169, 48.84 %	155, 43.42 %

TTX_7 Thermoproteus tenax: 1102723_1104458

GCGAAGAGGCGCAGGACAGTTAAAAAGCGTAAATCTATATATCCTAAATAATTCCCGGCGTTTAAAAATCGTCGT
GCAACTTGCCCTTTCTTTCTGTAAAATCGTTCAACTGCTGCGCCGTTGTACCTTTGTCAAGCGGGGATCCCGTTT
AAAAAACGGCGGACAGAAGTACGGCGGACGGCGGCGCCGACAGTAAAACACGCGCTTTCAACTGCCGGAGGCAGA
AGGCCAGCTGTGGAAAAGAGACAACCCCACCCCTTAGCCAACACGAGAACACGCCAGAACTAACACAAAACCCAC
CAGAGAAAAACTTAAAAACAGCCAAAAGCCAAAACCCAGAAGGCCCGGAATCTCAAAGAGAGGATTGAAAGTCAA
ATTCACAATCAACTTTCTACTGAACTTCCCGTCATGGAGAATCTCAAAGAGAGGATTGAAAGGCGCGCTGTATAC
AGCGCGCCTTTCGCGGCAGATATCGCCCTCAGCGAATCTCAAAGAGAGGATTGAAAGCTTCGTCGAGAAGTCCTC
GACGCTCAGAGACATATTCTACAACAAGAATCTCAAAGAGAGGATTGAAAGCTTCTTACTTCCTCAAGTCTCCTG
GCTATCTCCTGAATTTTTTCGAATCTCAAAGAGAGGATTGAAAGAGGCTACGCGATGTGCTAATTCAGTTGCGCG
ACGTGGCCGACTCGAATCTCAAAGAGAGGATTGAAAGTCGACGACGATGAAGTCCGGATTGCTCGCGCTTATCAG
CCTGAACGAATCTCAAAGAGAGGATTGAAAGGTAGATCTCGACTGCCACCGCCGGCTTGGCGCCGCGCCACGTGT
AAATTGAATCTCAAAGAGAGGATTGAAAGGGATAGACACGGCGGAAGCAATATGGACATCTGATCTCATCGAGAA
TCTCAAAGAGAGGATTGAAAGCTTTTATACATCTTTCTTTTTCGGCTATGTGCGACAACTGAATCTCAAAGAGAG
GATTGAAAGTCTAGCTCAACTACCTTTTCGCCCATGTCCAAAGGCCCAGACCTGAATCTCAAAGAGAGGATTGAA
AGAACAATAATGATATCGTGTGCAAAGATAGACAAAAGCGGCGAATCTCAAAGAGAGGATTGAAAGCTTGGGCGT
TGGAGCGGACAGAGCCCGCGCGCTCTTGAGGGAGAGGGAATCTCAAAGAGAGGATTGAAAGCTTCCTCACTGCGC
ACTCTCTCGTACTTGCACAGTTTGGCGAATCTCAAAGAGAGGATTGAAAGACCGCCAGCTTGAGCATGTGCGTAG
TCCCGCCGGTCTCCCACGGAATCTCAAAGAGAGGATTGAAAGTCTGGTGGAACCATAAACGGCAAAACCAATTCA
TCGTCCCAGACTTTGATTTTGTGAATCTCAAAGAGAGGATTGAAAGTCCATTGGCGCTTAGCCGGCCTTTGAAGC
GCTCTCCGCCTACACTTTAATCGATGGGCGGGGGCCTCGCCGGCAGGCATGCCCCCGGTACCGCATTACTTGGCT
CTGACAGTTTCTGGCAGTTGGGTGGCCGATAGAAGGACTATCGTAGAGGATATTAGGGCGGTCAACGAGAGATCT
CTCTCTAGGCCGAAAGGCAGCAGCAGGGCTAACGTGACTCCGAAGCCCCCTCCCAACAATCTGCCTAGAAATAAC
GTTATATTGGTCCCCGTGGCTTTGAAGTTGGCGGGGTAGACCTCGCTCATCCAGACCCCAAAGTAGGCGAAGAAG
GAGGAGCCGA

	Flanking site 1 (347 bp)	Flanking site 2 (339 bp)
g,c	173, 49.86 %	194, 57.23 %
a,t	174, 50.14 %	145, 42.77 %

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	Genome	bp	TTX genome, E-value 0.1	Archaeal viruses, E-value 0.01
TTX_1	Thermoproteus tenax: 227260-224864			
1.1	CCGGCGCCGACATTGATGCATATGTTGTTCACATCGAGATAGT	43	-	-
1.2	TTTCTCTTTCAATTTTTCTCCCTAATCTTCTCCCTAAACA	40	TTX_0712_iorA, indolepyruvate fd oxidoreductase, 16/16, 0.010	ATV_ORF209, 17/17, 1e-04 / SIRV2_Cas4, ORF207, 15/15, 9e-04
1.3	CGGCGCACAGTCTGTATCGTTCACCCCTGTGAGCGGCTATAC	42	-	-
1.4	TCGGTGATGGAGTACCTGTCTATGCCCGCCTCATCGAGCACTCTGG	46	-	-
1.5	CTTAGTATAGGCCAAATGACAACTAGCAGGAACACTATCAGCAAT	45	-	-
1.6	AACTTCTTCGTCGTGGAGGGGCCGGACGGCCGGCAATACAGG	42	-	-
1.7	TTGCCGGCGTCTGTCAGCACGGCCGTGGTTTCCCTCCAATT	41	-	-
1.8	TTGGACGCTGTGTTGGCATGCGCCGACCCGCGGCGGATTGCT	42	-	-
1.9	ATTGGATATATGTGCGGCATCCTCGTATTGCGAAATACGCTCCA	44	TTX_1000_asnS, asparaginyl tRNA synthetase, 16/16, 0.011	TTSV_hypothetical ORF16, 13/13, 0.009
1.10	GAAAGCCCAACCTAAGCTTCGTCTCAAGCATGGCGAGATCGT	42	-	-
1.11	TACCAGGTGCAGTGCAACGGCATCAGCTATACCGACTACACCT	43	-	-
1.12	CAACTGAGCATCACGCAGGGGGCGGTCGTCGACGGAGGGTACCTCTACTC	50	-	-
1.13	ACCCAGTAGGATATCGCCTCATAGCGGTATTTGAAATCGCC	41	-	-
1.14	TACGAATATATCGGACGCAAGGTCGTAAAGTTCGAGCTCTTCTG	44	-	-
1.15	TCTTTTGTCACTATCACAGCCAATCCCGAATCGAGACGGACC	42	-	SSVR_hypothetical ORF E152, 13/13, 0.007
1.16	CTTATCCTGACCGGCTTCGCGTACGGCTTGTTCACATATCTGCAC	45	-	-
1.17	TTTCTGAAGGAACGGGTGAAGGAGGTGCTGAAGAAGTACGGCG	43	-	-
1.18	GCGTAGCTGTTCCTCCGTTATCTTTATTGCGATGTTCTC	39	-	-
1.19	CCCTTGCTTAGCGAGCTCATTGGCGACAAGGTTGATCTTGTTAT	44	-	-
1.20	ATATTCGACGAGACGGGCAACACGAAGATCGGCACATGGATCT	43	-	-
1.21	CAGATGCGCCAAGCTATCGAGTCCACGTGTCCAACCGAGCG	41	-	-
1.22	TACGTGCGCGTTGAGCGCGATATGAAGTTCGACGCAGTCGAG	42	-	-
1.23	AGGAGGTCGCCGTCCCAAGATGTATTGCTTCGACTATGAGCAAG	44	-	-
1.24	GGGTTCCAGCCAAGCGCATGCGCCAATCCCCATCCAGCTATT	42	-	ARV_hypothetical gp26, 16/17, 0.01
TTX_2	Thermoproteus tenax: 316783-317791			
2.1	TGTACTTTAGGAATAATATACTCGTCAAGTATCTTGTAATACA	43	-	-
2.2	TTGCCCTTCAGAAGGTCGCCCTCCTTTACGTTGCAACCCCAA	42	-	MP_733-923 intergenic region, 13/13, 0.01
2.3	ATTAGACGTTGTAGCACTAACGCCTCTTTCACAACATTTATA	42	-	PSV_hypothetical ORF183, 14/14, 0.003
2.4	GCCGGGCATGTGAGCCCCCGTTCGATGTGCTCATAAGTCTGC	42	-	-
2.5	GATCAAAAAAGGAGGTGGACGTTCAACGTCCTCTGCACGTACCTAG	46	-	-
2.6	ATACACATCGGCACCCCATGCTATGTATTTTTCTCGCGTGGTA	43	-	SIFV_hypothetical 0032, 14/14, 0.004
TTX_3	Thermoproteus tenax: 346121-345073			
3.1	GCTTCCACAAAGTGAGCACCGCCAGTCTTAGAAATGTAGCCTAGCT	46	-	-
3.2	GGTTATCTCCGATGGACTGGCCACACACAGAGAGAGGTAGGTGC	44	-	-
3.3	GATCACGTTTGACTTCCACAGTTACGTGGTCGGAAGAAAAAGCGGGG	47	-	-
3.4	GTACGGGAGGGGATATTTTAGCTCCCCAGCCGCTCCAGAGACCTCTTCG	49	-	-
3.5	TGTGAGCCATGGCACTACAGGCCCCCACGCTGTATCGCCATTGA	44	-	-
3.6	TATCGCGATGCTATATAGCCTCGCGTGGACGGCGTAGAGCGCGT	44	-	-
3.7	CGTTCTAATTCTCGTCTTGATCCTCTCCTCTGCTCCTCAGGG	42	-	STSV_hypothetical ORF46, 14/14, 0.007

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	T. pendens / P. aerophilum / T. neutrophilus genomes, E-value 0.1	CRISPRdb, E-value 0.1	BLASTn program, word size 7, E-value 0.02
TTX_1	Thermoproteus tenax: 227260-224864		
1.1	-	-	-
1.2	-	Thermofilum pendens Hrk 5, 10.6, 20/22, 0.013	-
1.3	-	-	-
1.4	-	-	-
1.5	-	-	-
1.6	Alcohol dehydrogenase GroES, Tpen_1516, 19/19, 2e-04	-	-
1.7	-	-	-
1.8	-	-	-
1.9	-	-	-
1.10	-	-	-
1.11	-	-	-
1.12	FeS assembly protein SufB, Tpen_1206, 16/16, 0.014	-	-
1.13	-	-	-
1.14	-	-	-
1.15	-	-	-
1.16	-	-	-
1.17	Ribosomal L18P, Tpen_0237 / hypothetical, Tneu_1415, 16/16, 0.011	-	Trypanosoma cruzi strain CL Brener trans-sialidase partial mRNA, 25/25, 6e-04
1.18	Malate dehydrogenase, PAE2370, 17/17, 0.003	-	-
1.19	-	Wolinella succinogenes DSM 1740, 1.2, 15/15, 0.052	-
1.20	-	-	-
1.21	-	-	-
1.22	-	-	-
1.23	-	-	-
1.24	-	Pyrobaculum aerophilum str. IM2, 5.4, 14/14, 0.17	-
TTX_2	Thermoproteus tenax: 316783-317791		
2.1	-	-	-
2.2	-	-	-
2.3	-	-	-
2.4	-	-	-
2.5	-	-	-
2.6	-	Syntrophomonas wolfei, 2.29, 16/16, 0.014	-
TTX_3	Thermoproteus tenax: 346121-345073		
3.1	-	-	-
3.2	-	-	-
3.3	-	-	-
3.4	paREP2b, PAE0850, 17/17, 0.004	-	-
3.5	-	-	-
3.6	-	-	-
3.7	hypothetical, Tpen_0049, 18/18, 7e-04	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	Genome	bp	TTX genome, E-value 0.1	Archaeal viruses, E-value 0.01
TTX_4	Thermoproteus tenax: 1078562-1075638			
4.1	TGTATTGAGAGATGTAGTTGATTATTCGTGCGGCGCAAATGTCCGC	46	-	-
4.2	CTCGCCGCTCACGGGGAGGAGGAGGAAGAGGCCCACGTAGGCGTCGTATA	50	TTX_0193_hydrogenase maturation, 16/16, 0.014	-
4.3	AAGTACATCCGCGATCGTTAACCTCTTTTTTCTCATCTT	39	-	SIFV_36821-36854 intergenic region, 16/16, 2e-04
4.4	TGACACAAAGGGCGTTGACGTCACCGACATCCCCGACGC	39	-	-
4.5	TCTTCGCCCCGGCCGTCTGGGGAGCCGACTGGCTCCTTCCCA	42	-	-
4.6	CGTATTTAACATCACAGCTGTGTTTCAGAGCGCCTCA	37	TTX_0660_hypothetical, 37/37, 3e-15	-
4.7	CGAAGTCGGATATGCGATCGCATGCCTGACCACGTAGATCCGCGT	45	-	-
4.8	CTATATACACGTATATACACAGAGGGGCCCTTTGAAGGCGG	41	-	AFV7_putative posphatase gp52, 14/14, 0.004
4.9	TCGCGGGGGAGCGGATACAAGACATCCTGGCGTGGTACGTCGC	43	-	-
4.10	TGCTTAACCACTCGCCCGGTATTTCCTCGATCTTTTCTTC	40	TTX_0926_putative oxidoreductase, 16/16, 0.010	AFV6/7/8_16832-975/16330-474/15356-498 intergenic, 16/16, 2e-04
4.11	TCGAGTCGCTCTGCGGCGGTGAGCCGCTGTCCGAATGTGT	40	-	-
4.12	CCACTGCGCTATATGTTCCAGGACGTGCTGGAGGATGT	38	-	-
4.13	GCTCGTGTACACGCAGCGCAAGCACTCACAGTGTCTCGTCG	41	-	-
4.14	TAGAGTACCACCCAGACGGCCGCATGCTGAGCAAGTGCGACA	42	-	-
4.15	CCTTGCGGGGCACGCTCCTCAACCCCTCCGTCAGCCCCCTCCAGTCGAT	49	TTX_0673_conserved hypothetical, 16/16, 0.013	-
4.16	ACTGGCGGATACACGTTATGCCGATTGGAGAGTCAATATACTT	43	-	-
4.17	GCAGTTCGCCTTGTGGGCCTACTTGATAAACGTGTTCTTCACAG	44	-	-
4.18	TCTTGACTAGATAGCCCTTCTTGCGCTTTTTCTCCGATGGCT	42	-	-
4.19	AATCCCTGTACAGCTCCTCAATATAGCGCACATATTCGCGCCA	43	-	SSV5_hypothetical gp10, 13/13, 0.006
4.20	TTGAGAATATTCTATAAAGAATTGCTGAATTTTCTGATGGGC	42	-	TTV_38.6 kDa protein, 6625-6646, 18/22, 0.050
4.21	CGATTTGAGGAAAAAACCTTTAGGCGCGGCGGCTATGTACTAT	43	-	ATV_34623-34704 intergenic region, 14/14, 0.007
4.22	TCTTCTCGAGGTATCCCAATCTGCCAGCAAAGAGCTCGAATCT	43	-	-
4.23	CTTTTCCCTCATTCTGTCGACGAGGATCTCCTGGCTTTCTTCCTGG	46	-	MWP_phage portal protein ORF10, 14/14, 0.003
4.24	AGTCGTGTCCGCTACCACCTCCTAATCCCTTCTTTTTTCG	40	-	SSVR_10895-10932 intergenic region, 13/13, 0.006
4.25	GCGACGATCAACATCGGCACAATTAACATCGATACCATCGTGAAG	45	-	SIRV1_hypothetical ORF562 14/14, 0.004
4.26	ATCTATTGATGACGAAGCTAACAATATCAGCTTGAACACACCATTG	46	-	-
4.27	GGCCTGACGGACAGCACTGAGCCGTCAAAGAGGACGGACAAC	42	-	-
4.28	CATGTCACAGGATCATGCTGATAAGAGAAAACATGGAATATACG	44	-	-
4.29	GTAAGTGACAATCTTACAGAGTACTTCTCTATTGCTCTCCATAGAGCTTAAGCACTC	57	-	TTV_10 kDa protein, 2007-2036, 23/30, 0.075
4.30	TAAGACTTGGGCCACAAGTCTTAAGAATGAGGAATAGAA	39	-	AFV_hypothetical ORF65, 16/17, 0.008
4.31	TTTTCTCTACATAGCTTTTCTTTATCGTTATCTCAACACTGAACT	45	-	-
4.32	CAACTTCTTTTTCTGTATACTCTAATTTGTATGCAAGCGAT	41	-	AFV2_hypothetical gp02, 14/14, 0.003
4.33	TACCAGGAGTTCTGGCTACTCTACTTCCAGGCGCCGTTTTGG	42	-	-
TTX_5	Thermoproteus tenax: 1081524 1084343			
5.1	TCTGGACTGTGTGTTGCAACAGCGAAGTCGCGAACACCTCGAC	43	-	-
5.2	CGTTGATGTGGCCGGGACTGGCTGACTCAGCTATTACGTT	40	-	PSV_hypothetical ORF76b, 15/15, 7e-04
5.3	ACTTCAAGGCGTGCTTGCTGGCGTTTTAGGCGACGATGAGCT	42	-	TTSV_8826-9340 intergenic region, 13/13, 0.008
5.4	CTCTTTTTCCTGCCCCGTTCTTCCAACTCTTCGCCGGTCGT	41	-	-
5.5	CCGAGTAGCCGACGGCGCGTCCGCTCACTTAAGATCTCA	39	-	-
5.6	TGTGATAGATGAGATAGACAAAATCAAGTTCAGCACCACGT	41	TTX_0274 cell division control cdc21, 22/23, 2e-04	SSV4_hypothetical ORF73, 14/14, 0.002
5.7	GCGCACTAATTACCGCCCCGGCGCAGGATTACCAATCTTGCCGGGCCGGTT	51	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	T. pendens / P. aerophilum / T. neutrophilus genomes, E-value 0.1	CRISPRdb, E-value 0.1	BLASTn program, word size 7, E-value 0.02
TTX_4	Thermoproteus tenax: 1078562-1075638		
4.1	-	-	-
4.2	hypothetical, Tneu_0386, 17/17, 0.003	Thermococcus gammatolerans EJ3, 3.2, 16/16, 0.017	-
4.3	-	Thermomicrobium roseum DSM 5159, 1.61, 17/18, 0.042	-
4.4	-	-	-
4.5	-	Streptomyces avermitilis MA-4680, 12.26, 17/17, 0.004	-
4.6	-	-	-
4.7	-	-	-
4.8	-	-	-
4.9	-	-	-
4.10	-	-	-
4.11	-	-	-
4.12	-	-	-
4.13	-	-	-
4.14	-	-	-
4.15	-	-	-
4.16	-	-	-
4.17	-	-	-
4.18	-	Streptococcus pyogenes M1 GAS, 1.1, 15/15, 0.048	-
4.19	-	-	-
4.20	-	Synechococcus sp. PCC 7002, 1.12, 16/16, 0.014	-
4.21	-	-	-
4.22	-	-	-
4.23	-	-	-
4.24	-	-	-
4.25	-	-	-
4.26	-	-	-
4.27	-	-	-
4.28	-	-	-
4.29	heat shock protein Hsp20, Tneu_1620, 16/16, 0.016	-	-
4.30	-	-	-
4.31	-	-	-
4.32	-	-	-
4.33	-	-	-
TTX_5	Thermoproteus tenax: 1081524 1084343		
5.1	-	-	-
5.2	-	-	-
5.3	-	-	-
5.4	-	-	-
5.5	-	-	-
5.6	-	Clostridium novyi NT, 3.22, 18/19, 0.013	-
5.7	-	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	Genome	bp	TTX genome, E-value 0.1	Archaeal viruses, E-value 0.01
TTX_5	Thermoproteus tenax: 1081524 1084343			
5.8	TTGGTCTCAGGCGCGCGGTAGCCTTCCAGCTGTACGACCAACTCGAT	47	-	-
5.9	TCCGGAGGAGGCGGCGAAGAGGATGATAGAGTTGGCCAGGCTGG	44	TTX_0331_hypothetical, 19/19, 2e-04	TP1_peptidoglycan hydrolase p108, 14/14, 0.009
5.10	GGCATGATTGGTTGCGCCGGGAGCTCTGGGATGTGATCCGCCA	43	-	-
5.11	CTACTTGGAGTATGCCGGTCAAATTCGGTGGGGCGGACGTCGTACC	46	-	STIV_hypothetical B130, 13/13, 0.008
5.12	TTTGTTGAAATAATCACCGTAACCCAGGCTTTCCAACCATCTT	43	-	STIV_9049-9392 intergenic region, 13/13, 0.007
5.13	ATCGCTTGTCCAACCGGGCTCCTCTATAATTGTCGTCATTAGCTTA	46	-	-
5.14	ACTGGATCATCGATGGGCAGAATGCGGGCGTGTCGGCTATATG	43	-	-
5.15	CAAGTCGTCGAGAAGTATTTTGGTCTGCTGAAGCAGTACGGTTTCG	46	-	ARV_glycosyl transferase gp23, 15/15, 7e-04
5.16	GTACCACGACATCAGCCATGTGACTGTTGCGCCTAGCTCAATCTC	45	-	-
5.17	ACTTGAATTTCAACAATGTCACCAGCACATATGTCCAGGCTAGG	44	-	TTV_Full coat protein TP1, 1773-1800, 23/28, 0.004
5.18	ATTTTTATTCCTCATAAACATGGTCTCAAGGTCAGAACTAAGCC	44	-	SSV2_hypothetical ORF276, 13/13, 0.006
5.19	AATGGCTCCAAGGGATCGTTCTCTAGACTCCAATTTTTCTAAT	43	-	ATV_hypothetical ORF100, 14/14, 0.007
5.20	TTTGTCGACGGGATCTTGCGACATGCGGTACGCTGCAGTGTCG	43	-	-
5.21	GCTCCACCGACCAGTACCCATATAGCGTGAACTCTTTCGCCGGA	44	TTX_5.26, 19/20, 0.011	-
5.22	AGTGCGCAGACCCCTGCACAGGCCGGCGTCCTTCACAAAA	40	-	-
5.23	ATGTAGAGCTTGACCGGCGCCCCAGGATAGGGCACCGGGCCG	42	-	-
5.24	TCTGTATCCCGCCGGCGCGTGGAGAAGGCGTTTCTCCACAACGTCGTCCTC	51	-	ABJV_hypothetical gp22, 14/14, 0.005
5.25	TCGAGGACGATGTAGTCGGGATTGCTCGCGTTTATCAGCTTGA	43	TTX_7.6, 35/39, 7e-07	-
5.26	CTCGTGGATACTCCAAGCGAAAGAGTTCACGTTATACGGCTAC	43	TTX_5.21, 19/20, 0.011	-
5.27	TCCACAAGCCTCACGGCCGTCTTCACGGCGCGGCCCGGTCTGACGG	46	-	-
5.28	GGGATATACTACACGGCCCTGCGCGTCATCGGGAACAAGCGCATGG	46	-	SIRV2_hypothetical ORF110, 14/14, 0.004
5.29	TATTCCGTAGATTCCGGCGTAAAAGGTCTGGGCGGCGAT	39	-	TTSV_20707-20933 intergenic region, 13/13, 0.008
5.30	CCGTGCACCTATGACTTTCAAGTACTCAAGGACGTTAG	38	-	-
5.31	AACGCTTGGACAAATACCGTACCGCTACCCTCATATTGATA	41	-	PSV_hypothetical ORF101, 14/14, 0.003
TTX_6	Thermoproteus tenax: 1097343-1094882			
6.1	ATAGCTATCTAAGCGCAATAGTAGAAATTCCAATTCTAAGCTTT	44	-	-
6.2	AGCTCTGGGTATTTGACAATTACTTCGGGCTTGGGGATGCTCGTG	45	-	-
6.3	CCTCGTTCCTTTTCGCTATCACTACATCTCCTAATTTGTTT	41	-	AFV9_glycosyl transferase gp58, 14/14, 0.004
6.4	CCGCAAGGCCAGTGCATGGATTCTCCGTACCTCTCAGTGAACC	43	-	AFV2_glycosyl transferase gp22, 14/14, 0.003
6.5	TCAGAGCAAGGGGTACAAGGTCTACCAGTACGGGGACGGTCTCATAACAGT	51	-	-
6.6	TCGCAGTTCTCTATTGTTATTAAGGCCTTATAATTTGTATATGTTCCCT	49	-	TTV_7.5 kDa, 1131-1146, 15/16, 0.018
6.7	GGGAATACATTAAGACCAATAAAAATTGTTCACCTCGGCTTT	42	-	TTV_26.8 kDa, 6182-6204, 21/24, 0.004
6.8	TTGAACGTCACGCCGGCGCTGTCTCAACTACTTCGCGGTGGCT	43	-	-
6.9	TTCATCGTGGCGGCCTTCCTATTCTGGTGGTGGCTGAAGAA	41	-	-
6.10	GGGACAGGTTCAAGGGGTTCCTCGGCCTCCTCATCGTGCT	40	-	TP1_tape tail measure p96, 15/15, 0.002
6.11	GTTGACCAGGTACTTCGACAACCTTTACACAGCCAGG	37	-	MP_hypothetical ORF1, 13/13, 0.009
6.12	GCCGGCCTTCGTCGCATGCTCCGACGAGTGGAGCCACGCT	40	-	AFV6_hypothetical gp31 / AFV9_hypothetical gp28, 16/16, 2e-04
6.13	ACCTGCATCTCCGACGTTGGGGATGACGCGGGCGTCG	37	-	-
6.14	TAATGGATCAATATAGAACTCAGTCTCGTGGTTGGTCTCATATA	44	-	-
6.15	CAGGGCATAAAGCCTATTGCGTTGAAGTGCTTAACTACCTCTT	43	-	-
6.16	ATCTCCGCCTTTATCGCCAGAATTCTGCTGTCCATTTGGGTTTCGGG	47	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	T. pendens / P. aerophilum / T. neutrophilus genomes, E-value 0.1	CRISPRdb, E-value 0.1	BLASTn program, word size 7, E-value 0.02
TTX_5	Thermoproteus tenax: 1081524 1084343		
5.8	hypothetical, PAE0356, 17/17, 0.004	-	-
5.9	intergenic Tpen, 418796-79, 18/18, 7e-04 / metallophosphoesterase Tneu_1820, 17/17, 0.003	-	-
5.10	-	Pyrococcus furiosus DSM 3638, 2.14, 19/21, 0.17	-
5.11	-	Chloroflexus aggregans DSM 9485, 36.59, 15/15, 0.055	-
5.12	-	-	-
5.13	-	-	-
5.14	-	-	-
5.15	-	-	-
5.16	-	-	-
5.17	-	-	-
5.18	-	-	-
5.19	-	-	-
5.20	-	-	-
5.21	-	-	-
5.22	CRISPR 6.6, Tneu, 17/17, 0.002	-	-
5.23	-	-	-
5.24	-	Pyrococcus furiosus DSM 3638, 8.4, 15/15, 0.06	-
5.25	-	-	-
5.26	-	-	-
5.27	hypothetical, PAE1836, 17/17, 0.004	Frankia alni ACN14a, 2.18+2.34, 15/15, 0.055	-
5.28	-	-	-
5.29	-	-	Pseudomonas putida KT2440, outer membrane OprG, 23/23, 0.007
5.30	-	-	-
5.31	-	-	-
TTX_6	Thermoproteus tenax: 1097343-1094882		
6.1	-	-	-
6.2	-	-	-
6.3	transcriptional regulator XRE, Tneu_0689, 18/19, 0.039	-	-
6.4	-	-	-
6.5	-	-	-
6.6	-	-	-
6.7	-	-	-
6.8	-	-	-
6.9	-	-	-
6.10	-	-	-
6.11	-	-	-
6.12	-	-	-
6.13	-	-	-
6.14	-	-	-
6.15	-	-	-
6.16	-	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	Genome	bp	TTX genome, E-value 0.1	Archaeal viruses, E-value 0.01
TTX_6	Thermoproteus tenax: 1097343-1094882			
6.17	ATGAGATCCCACCAGTTGAACGTGCCCGCCTGGCTCG	37	-	-
6.18	TATCGACAAAGCATCTAATGCCACGCTCAACGATATGTATCTC	43	-	-
6.19	ATCAGCCGGCTCAAGTCGTTTGGTACGACAACGTAACGATA	41	-	-
6.20	TGAACTTTCTACAAATACCTGCATTATATTATCCGTAGAACCTA	44	-	SIRV1/2_Cas4 exonuclease ORF207/gp19, 14/14, 0.003
6.21	CGACAGCCTGACGCGATGATGGCCGCGGGGAGCCCGGCG	39	-	-
6.22	GCGGAGAACGAGTACGTGATAGTTGTCCCGGCCGGAAGTAGCACG	45	-	PAV_concanavalin A-like lectin/glucanase, 13/13, 0.008
6.23	ATTTATTTTTATTCTTTTCTGGATCTGAGTTTGGGCCTTCA	41	1289196-1289212_intergenic region, 17/17, 0.003	SSV2/4/5_zinc finger protein ORF 155/124/gp31, 13/13, 0.006
6.24	ACCCAAGTACAAAGTCAATGAGGCCGGGGAAATTGAGCTCG	41	-	PAV_concanavalin A-like lectin/glucanase, 13/13, 0.007
6.25	CGATGGGCGACCTCGAGACGAACTCCTCAAGGACGCGCCTGACCGGC	47	-	PAV_leucine zipper winged helix DNA-binding domain, 13/13, 0.009
6.26	CTTCAATCTCATCTAGTGCGTCCTCAAGCCGCATCAGCACATATA	45	-	-
TTX_7	Thermoproteus tenax: 1102723 1104458			
7.1	TCAAATTCACAATCAACTTTCTACTGAACTTCCCGTCATGGA	42	-	-
7.2	GCGCGCTGTATACAGCGCGCCTTTCGCGGCAGATATCGCCCTCAGC	46	reverse complementary 7.2 + repeat part, 25/25, 5e-08	-
7.3	CTTCGTCGAGAAGTCCTCGACGCTCAGAGACATATTCTACAACAA	45	-	-
7.4	CTTCTTACTTCCTCAAGTCTCCTGGCTATCTCCTGAATTTTTTC	44	-	-
7.5	AGGCTACGCGATGTGCTAATTCAGTTGCGCGACGTGGCCGACTC	44	-	SSV4_hypothetical ORF152, 13/13, 0.007
7.6	TCGACGACGATGAAGTCCGGATTGCTCGCGCTTATCAGCCTGAAC	45	TTX_5.25, 35/39, 8e-07	-
7.7	GTAGATCTCGACTGCCACCGCCGGCTTGGCGCCGCGCCACGTGTAAATT	49	-	-
7.8	GGATAGACACGGCGGAAGCAATATGGACATCTGATCTCATCGA	43	-	-
7.9	CTTTTATACATCTTTCTTTTTCGGCTATGTGCGACAACT	39	-	ABV_hypothetical ORF462, 13/13, 0.009
7.10	TCTAGCTCAACTACCTTTTCGCCCATGTCCAAAGGCCCAGACCT	44	-	AFV9_viral structural protein gp33, 14/14, 0.004
7.11	AACAATAATGATATCGTGTGCAAAGATAGACAAAAGCGGC	40	-	SIFV_superfamily II helicase 0007, 14/14, 0.004
7.12	CTTGGGCGTTGGAGCGGACAGAGCCCGCGCGCTCTTGAGGGAGAGG	46	-	-
7.13	CTTCCTCACTGCGCACTCTCTCGTACTTGCACAGTTTGGC	40	-	-
7.14	ACCGCCAGCTTGAGCATGTGCGTAGTCCCGCCGGTCTCCCACG	43	-	-
7.15	TCTGGTGGAACCATAAACGGCAAAACCAATTCATCGTCCCAGACTTTGATTTTGT	55	-	-

A2: Nucleotide sequences of CRISPR spacer and results of similarity searches

Cluster	T. pendens / P. aerophilum / T. neutrophilus genomes, E-value 0.1	CRISPRdb, E-value 0.1	BLASTn program, word size 7, E-value 0.02
TTX_6	Thermoproteus tenax: 1097343-1094882		
6.17	-	-	-
6.18	-	-	-
6.19	GAPDH, Tpen_0757, 18/19, 0.040	-	-
6.20	-	-	-
6.21	transferase, Tpen_0895, 21/23, 0.037	-	-
6.22	-	-	-
6.23	-	Sulfolobus tokodaii str. 7, 2.24, Sulfolobus solfataricus P2, 5.27, 17/18, 0.046	Yarrowia lipolytica strain CLIB122, YALI0B17028p, 23/23, 0.008
6.24	-	Geobacter sulfurreducens PCA, DR1, 15/15, 0.001	-
6.25	-	-	-
6.26	-	-	-
TTX_7	Thermoproteus tenax: 1102723 1104458		
7.1	-	-	-
7.2	hypothetical, Tneu_1627, 17/17, 0.003	-	-
7.3	-	Thermococcus onnurineus NA1, 2.5, 21/24, 0.19	-
7.4	-	-	-
7.5	-	Frankia sp. EAN1pec, 7.31, 17/17, 0.004	-
7.6	-	-	-
7.7	hypothetical, Tpen_1858, 16/16, 0.013	-	-
7.8	-	-	-
7.9	-	-	-
7.10	-	-	-
7.11	-	-	-
7.12	-	Gluconacetobacter diazotrophicus PAl 5, 1.11, 17/18, 0.055	-
7.13	-	Hyperthermus butylicus DSM 5456, 2.21, 19/21, 0.044	-
7.14	-	-	-
7.15	-	-	-

A3: Detailed identification of *T. tenax cas* genes. ORF ID, classification (according to Haft *et al.*, 2005), COG and TIGR numbers

gene-Nr	gene sequence	family	subfamily	COG	TIGR	name	strand	length, aa	comments
TTX_0232	219442-220191	RAMP	1583	5551	TIGR01877	Cas6	-	249	pfam10040, DUF2276
TTX_0232a	220188-220766	2254	2254	-	TIGR01596	Cas3_HD	-	192	HD superfamily hydrolase, possibly a nuclease
TTX_0233	220763-222553	1203	1203	1203	TIGR01587	Cas3	-	596	cd00046, DEXDc, DEAD-like helicases superfamily, Mg/ATP-binding
TTX_0234	223147-223929	RAMP	1688	-	TIGR01874	Cas5	-	260	-
TTX_0235	223926-224861	1857	1857	1857	TIGR02583	DevR_archaea, Csa2	-	311	pfam01905, DUF73, Cas autoregulator protein, Csa2 family
CRISPR TTX_1	224864-227260								
TTX_1228	1066926-1068347	1517	1517	1517	TIGR01897	Cas_Csx1	+	473	pfam09455, Cas_MJ1666 family, DxTHG-motif
TTX_1229	1068463-1068843	-	-	-	-	hypothetical protein	-	126	-
TTX_1230	1068840-1071089	1353	1353	1353	TIGR02578	Cas_TM1811, Csm1	-	749	Csm1 family
TTX_1231	1071076-1072071	RAMP	1332	1332	TIGR01899	Cas_TM1807, Csm5	-	331	Csm5 family
TTX_1232	1072075-1072893	RAMP	1567	1567	TIGR01903	Cas_TM1808, Csm4	-	272	Csm4 family
TTX_1233	1072890-1073813	RAMP	1337	1337	TIGR02582	Cas_TM1809, Csm3	-	307	pfam03787, Csm3 family
TTX_1234	1073810-1074235	-	-	-	-	hypothetical protein	-	141	-
TTX_1235	1074409-1074873	-	-	-	-	hypothetical protein	+	154	-
TTX_1236	1074912-1075202	-	-	3593	-	conserved protein	+	96	ATP-dependent endonuclease, OLD family
CRISPR TTX_4	1075638-1078562								
TTX_1240	1078563-1079240	1517	4006	4006	TIGR02619	Cas_Csx1	+	225	pfam0965, DxTHG-motif
TTX_1241	1079278-1079475	-	-	4938	-	conserved protein	+	65	MUTSac, ATPase domain of DNA mismatch repair MUTS family
TTX_1242	1079475-1079645	-	-	-	-	hypothetical protein	+	56	-
TTX_1243	1080237-1081523	1517	1517	1517	TIGR01897	Cas_Csx1	-	428	pfam09455, Cas_MJ1666 family, DxTHG-motif
CRISPR TTX_5	1081524-1084202								
TTX_1245	1084203-1084775	1468	1468	1468	TIGR00372	Cas4	-	190	pfam01930, DUF83, RecB family exonuclease
TTX_1246	1084759-1085553	1518/1343	1518/1343	1518/1343	TIGR00287/1573	Cas1 / Cas2	-	264	pfam01867, DUF48, pfam02647, DUF196, uncharacterized ACR
TTX_1248	1085643-1086479	1468	4343	4343	TIGR01896	Cas_AF1879, Csa1	-	278	pfam06023, DUF911, Csa1 family
TTX_1249	1086553-1087167	HTH	2462	-	TIGR01884	cas_HTH, Csa3	+	204	CRISPR locus-related DNA-binding protein, Csa3 family
TTX_1250	1087204-1087596	-	-	-	-	Csa5	+	130	-
TTX_1251	1087593-1088582	1857	1857	1857	TIGR02583	DevR_archaea, Csa2	+	329	pfam01905, DUF73, Cas autoregulator protein, Csa2 family
TTX_1252	1088582-1089262	RAMP	1688	1688	TIGR01874	Cas5a	+	226	pfam09705
TTX_1253	1089252-1090952	1203	1203	1203	TIGR01587	Cas3	+	566	cd00079 HELICc, smart00487 DEXDc DEAD-like, pfam00270 DEAD, ATP-binding
TTX_1254	1090949-1091629	2254	2254	2254	TIGR01596	Cas3_HD	+	226	HD superfamily hydrolase, possibly a nuclease
TTX_1255	1091626-1092711	-	Q57830	-	TIGR01914	Csa4	+	361	pfam09703
TTX_1256	1092781-1093578	-	-	-	-	hypothetical protein	+	265	putative ABC transporter, ATP-binding protein
TTX_1257	1093575-1094015	-	-	-	-	hypothetical protein	+	146	-
TTX_1258	1094079-1094732	-	-	-	-	hypothetical protein	-	217	-
CRISPR TTX_6	1094882-1097343								

A4: Nucleotide- and corresponding amino acid sequences of the *T. tenax* operons *casa1* (TTX_1248-1245) and *casa2* (TTX_1250-1255) and the Csa3 protein (TTX_1249). The operon *casa1* encodes the proteins: Csa1 (brown), Cas1/2 (blue) and Cas4 (green); the operon *casa2* is coding for: Csa5 (grey), Csa2 (green), Cas5a (pink), Cas3 (yellow), Cas3HD (green) and Csa4 (black). Start and stop codons marked in red and amino acid sequences colored.

casa1

ttcactacacctcatcaaaacttc**gtg**ttgacactcctggagatcgcccgccttctgagg
V L T L L E I A R L L R
agggccaaggtcacccagggccccgccgaggtttcgccggagctgaggggctgggcctac
R A K V T Q G P A E V S P E L R G W A Y
gacagacagccggtcaaacccccgcctacctaggcctcgccctctccgacttcgcctac
D R Q P V K P P A Y L G L A L S D F A Y
ggctactgccccacggggaggagcctctacctcaagtacgtgttgggggagaggccccag
G Y C P T G R S L Y L K Y V L G E R P Q
cccaccaagccgctcgccgaggggcaggccctacacgcggtcctcttcaaggcgctggag
P T K P L A E G Q A L H A V L F K A L E
gacttcaggaggtacgtatactcgggcgctcccatgtctcctcccggcgagggcatgccg
D F R R Y V Y S G A P M S P P G E G M P
gaggacctcagagctaaggccgaggcgctgtataggtacatcgccgtaaggctcaccggc
E D L R A K A E A L Y R Y I A V R L T G
gagtaccaatacgtgctggcctcccgcctggccagaagccgcgacgccgccgccttctac
E Y Q Y V L A S R L A R S R D A A A F Y
gccgcccccatagccgcccagatagctgtggacggcgccccctcggcctcagctacgtc
A A P I A A Q I A V D G A P L G L S Y V
gtggcggacggagtggcgctgggggccgtggtggagttcaagttcggccccagccagaac
V A D G V A L G A V V E F K F G P S Q N
gtggacgtagcgctggctgggtacgccatggccatagaggcggagtacggcgtccccata
V D V A L A G Y A M A I E A E Y G V P I
gactacggcatacacgtccagatagccgtggacgggcaggtggagtacagagcgtctgcc
D Y G I H V Q I A V D G Q V E Y R A S A
tacgtcttaggcgacgcgccccgcgccaagttcctcgaggccagagacgaggccattgac
Y V L G D A P R A K F L E A R D E A I D
gtcgtcgcctccgccagagacccaggccccgccccccagtgccccaagacgtgccctac
V V A S A R D P G P A P Q C P K T C P Y
tacagcgtatgcaggtcg**tag**tggcctcatacggcgtcaagataagggctaggaagggcc
Y S V C R S –
tcctcctcgtggagacaaaggagggcgccaggtagtaccctctgcacgag**gtg**gacgaa
V D E
gtcctcctactcaccggcggcatatccataacgacgagggcgctcagggccctcctcgcc
V L L L T G G I S I T T R A L R A L L A
accggggccacagtcgccgtcttcagcccccgcggagagccgttgggcatattcatgagg
T G A T V A V F S P R G E P L G I F M R
ccggtcggagacgccacgggggccaagaggaggtgccagtacaaggcggcggaggacggc
P V G D A T G A K R R C Q Y K A A E D G
agagggctacagtacgccaaaagttgggtcttcaagaagatattgggccagagggacaac
R G L Q Y A K S W V F K K I L G Q R D N
atcaaggcctggcgccgccgcctcagaggctacagccaatacgccgagtctctggccaag
I K A W R R R L R G Y S Q Y A E S L A K
gccctccctggcgctggacttcatggagcaatggagaccccgcgtagacgccgtcgtggc
A L P G A G L H G A M E T P R R R R R G

caagatggcggacaggctggagtccgagggcggcctactcacccgccggtcccgcctgga
Q D G G Q A G V R G R P T H P P V P P G
gctggccgccgcagtcctggaggagctccacgcggccaagaggcctctctccgccgagat
A G R R S P G G A P R G Q E A S L R R D
ccacagagaggtcagagctctggcgcgctccatatgtacgtaatagtggtatacgacata
P Q R G Q S S G A L H M Y V I V V Y D I
acggagaacgacgtcagagccaaggtggccgacatcctcagggcctacggcctagcacgc
T E N D V R A K V A D I L R A Y G L A R
atccagcgctccgcctacgtcggcaggctcccgccggccctcgtcaaggagctcgccgaa
I Q R S A Y V G R L P P A L V K E L A E
aggctcgcgagggccgtaagaggcgcgaacgccgacatagccatattcaaggtagacaag
R L A R A V R G A N A D I A I F K V D K
aggacgatagacacctccctgaggatacccccagacctcccgcgggac**atgtcgctctgc**
R T I D T S L R I P P R P P A G HM VS AL LC
cattaacccgagcctaatacgtcagtacctctactgccccatggccgcctactacatc
HI N P S L I R Q Y L Y C P M A A Y Y I
gccgcgggcgccccagagccccccaccctccggatgcagaggggcagagagatccagcag
A A G A P E P P T L R M Q R G R E I Q Q
gaggccgcacaagccgccgccaaagccctgggcgccgagagggccgaatactccgtccac
E A A Q A A A K A L G A E R A E Y S V H
ataaccgccccgcccttatgcggcacagtggacgccgtactctggatcaacggaaggccg
I T A P P L C G T V D A V L W I N G R P
agcccctcgaggtgaaggccgcggcgaggccccgccgcatccccatacaccacaaggcg
S P L E V K A A A R P R R I P I H H K A
caagccgccgcctacatcgccatggtgcagagggcgtacggaagggctgtcgccacagcc
Q A A A Y I A M V Q R A Y G R A V A T A
tatatatactacgctgagagcggccagatagcccagatacgcatggccaaagacctacag
Y I Y Y A E S G Q I A Q I R M A K D L Q
gagctcctcaactacgcagtctcgaggctacaacagatactccaggggaagccgccggtc
E L L N Y A V S R L Q Q I L Q G K P P V
ctaaacccaaatccggcgaaatgccaaaactgttggtacagaaaatggtgcagccacctg
L N P N P A K C Q N C W Y R K W C S H L
ccaacaacagtggaaaaaatc**tag**
P T T V E K I -

csa3

atctcgggttgcaggtggcttgca**atg**aggatgggcatcacggtggggtttgagcacagg
M R M G I T V G F E H R
ttcgcggtgcgggctctcgccaagctggggccggccgatgagtacatctttctctacgcc
F A V R A L A K L G P A D E Y I F L Y A
tacaccggcagagatgaggactcgaggtcggagagcaccgtgagggagctgatagccgcg
Y T G R D E D S R S E S T V R E L I A A
ttcgggaggggggcgggattcagggtggatctgaccaagcccagcgaggcgttggagcag
F G R G A G F R V D L T K P S E A L E Q
atctacggcctggggctggagagggtcgtgttggcgggcggtcctaggttgctggtgctc
I Y G L G L E R V V L A G G P R L L V L
ctcgctctattggcctccctcctccggagatccgaggtatatgtcctcccggagtactcc
L A L L A S L L R R S E V Y V L P E Y S
gcggagcccctggatttgaccccctcaccgcgttatgcgacctctgcaggctcagcccg
A E P L D L T P L T A L C D L C R L S P

```
gcgaagctgggcatagtggcctcagtcagcgggagggtggggatagacgacgtggcgcag
 A  K  L  G  I  V  A  S  V  S  G  R  V  G  I  D  D  V  A  Q
ttggtgggccgggactcgacgactgtggcgaggtacctcaaggacctaagcgaggcgggg
 L  V  G  R  D  S  T  T  V  A  R  Y  L  K  D  L  S  E  A  G
attgtgaagtccctcggcggcaggaggagggagtacttcgtcgacccgctcctcaccaca
 I  V  K  S  L  G  G  R  R  R  E  Y  F  V  D  P  L  L  T  T
gttgctcgtcttatgcaaaacagagctggatctttgtagacagtaattcttaataattat
 V  A  R  L  M  Q  N  R  A  G  S  L  -
```

casa2

```
aataattattgtattctcttcccc atg gagtctgtacaactcaagaaggaagagctagtg
                          M  E  S  V  Q  L  K  K  E  E  L  V
gagcgggccgcccgctccgtaaggccggccgtacatctggagatggcctacgatgtgttg
 E  R  A  A  R  S  V  R  P  A  V  H  L  E  M  A  Y  D  V  L
gacgagctctccagatcccctgagaagtacccggagcagttggcaaagctctcgaggatc
 D  E  L  S  R  S  P  E  K  Y  P  E  Q  L  A  K  L  S  R  I
gttgtgaaggtgctgaacgacgtagaggatgagctcgaacacaatccccagaacgaggaa
 V  V  K  V  L  N  D  V  E  D  E  L  E  H  N  P  Q  N  E  E
ttacagaaggccaggaataggctggcggcttggggcggatacgtggccgagctcgccaag
 L  Q  K  A  R  N  R  L  A  A  W  G  G  Y  V  A  E  L  A  K
cgccttgaggaggcggatgacagagaggaggatccggatggtgaggctcttctgcgccatg
 R  L  E  E  A  D  D  R  E  R  I  R  M  V  R  L  F  C  A  M
gccctagctcccgacaaactgactgtagagctgaagaagctgcttaaggggcg atga gg
 A  L  A  P  D  K  L  T  V  E  L  K  K  L  L  K  G  R  M  R
gtggctccccccctacgttagggccgcgggccgcttcgaggcgcagctgtcggtcttgacc
 V  A  P  P  Y  V  R  A  A  G  R  F  E  A  Q  L  S  V  L  T
ggcgccggcaacatgggcaactacaacatgcacgccgtggccaaggtgattcacggcggg
 G  A  G  N  M  G  N  Y  N  M  H  A  V  A  K  V  I  H  G  G
aaagcctacgaagtgcccgtgttgactgggaacgctctgaagcactggcactcggtgtat
 K  A  Y  E  V  P  V  L  T  G  N  A  L  K  H  W  H  S  V  Y
ctggccgaggtgtaccaagacctcgggggacgcagctgaacgaattctgcaagaagggg
 L  A  E  V  Y  Q  D  L  G  G  T  Q  L  N  E  F  C  K  K  G
gtggggctgaggggaagaacatgagcggcgaggacgcctccagcgaggccgaggccatt
 V  G  L  R  G  K  N  M  S  G  E  D  A  S  S  E  A  E  A  I
aaagacctctgcaacgacctccacggcttcctcctccctgacaagcagatcaagcgcgac
 K  D  L  C  N  D  L  H  G  F  L  L  P  D  K  Q  I  K  R  D
agcctcgtgagggtgtccttcgcggtgccggtgctggaggagaagaacctcgaggcggcg
 S  L  V  R  V  S  F  A  V  P  V  L  E  E  K  N  L  E  A  A
agcaaattcgccgtggtccacaacagggtggacccctcaagaggacgcagcaggtgaag
 S  K  F  A  V  V  H  N  R  V  D  P  F  K  R  T  Q  Q  V  K
tccaaggaggagcaggaggggacggagatgatggtcttcaaacaggagtactcatcggcg
 S  K  E  E  Q  E  G  T  E  M  M  V  F  K  Q  E  Y  S  S  A
gtctacggcttcgcggcggccatggacctcgggctctccggagtgccgctctaccaggag
 V  Y  G  F  A  A  A  M  D  L  G  L  S  G  V  P  L  Y  Q  E
ggagaaagtggtgtagacgatagggagagagcgctgaggataaagtcggcgcttgtggct
 G  E  S  G  V  D  D  R  E  R  A  L  R  I  K  S  A  L  V  A
ctgctacggctgttcaacggagttggctctaagcaggcgcgcgccctccccatcgcaagg
 L  L  R  L  F  N  G  V  G  S  K  Q  A  R  A  L  P  I  A  R
cttagggagctggtgatcgcaatatcggactcgcccatccccaacctcgtacacggcgcg
 L  R  E  L  V  I  A  I  S  D  S  P  I  P  N  L  V  H  G  A
```

```
tatccagactacgtcgcaaggtcggcggagctcctgacggcgtatttaaaggctgtgggg
 Y  P  D  Y  V  A  R  S  A  E  L  L  T  A  Y  L  K  A  V  G
aagaaggggtcgtgctgtgctacggcgtcgactgcccgcagagcaacaacgtcctcgag
 K  K  G  V  V  L  C  Y  G  V  D  C  P  Q  S  N  N  V  L  E
tgtaaaaaggccgacacgttggacgacctatttcaagaggcccttaggagggctctgccg
 C  K  K  A  D  T  L  D  D  L  F  Q  E  A  L  R  R  A  L  P
ggggggcaacccgctaggaggtgatgcactactacttgatcgaggccagggcgcctctg
 G  G  Q  P  A  R  R  -  M  H  Y  Y  L  I  E  A  R  A  P  L
tactcggtgaaggtggtggacgtctaccaggtggcctccgcatatccgctccccacgccg
 Y  S  V  K  V  V  D  V  Y  Q  V  A  S  A  Y  P  L  P  T  P
cacgccgtcgtgggggcgctgggggcggccatggcccacgccggcatgtgccggggtttg
 H  A  V  V  G  A  L  G  A  A  M  A  H  A  G  M  C  R  G  L
cagtgtatgaaggaggcggaggggctggtcgtggcggctaggccggcggctgtgggggag
 Q  C  M  K  E  A  E  G  L  V  V  A  A  R  P  A  A  V  G  E
ctggccaagttccccgcggtgttgtggaggacgcgggggtgctcgaggacaagtccctc
 L  A  K  F  P  A  V  L  W  R  T  R  G  V  L  E  D  K  S  L
ccggcgggccttgaggactacgcgggcgcccgggacgctatggttagggagtacgtctat
 P  A  G  L  E  D  Y  A  G  A  R  D  A  M  V  R  E  Y  V  Y
gcccacgccgtcaagatactgctggtcaccaggaagagagtgccgcaacacgtgattagg
 A  H  A  V  K  I  L  L  V  T  R  K  R  V  P  Q  H  V  I  R
cttatggcgagactcggcgacagcgagtcctacgtgtcagtcgtcgacgtcttggagggg
 L  M  A  R  L  G  D  S  E  S  Y  V  S  V  V  D  V  L  E  G
aagccgagggagtgcgggggggcttgtcaacgtcgccgtgagggcggcgaaggcgaggaga
 K  P  R  E  C  G  G  L  V  N  V  A  V  R  A  A  K  A  R  R
ggtagctacacgctctaccgcgccttagacgagaggggaaccgggagctcttcgcatac
 G  S  Y  T  L  Y  R  A  L  D  E  R  G  N  R  E  L  F  A  Y
cccgtggtggaggagggcggggtctacaggcgcggcggcgtggaggttgggtctaaggtg
 P  V  V  E  E  G  G  V  Y  R  R  G  G  V  E  V  G  S  K  V
ctctgcttaggcgaggctgtgttccctgagggagatggttggtagaaggcttgacacg
 L  C  L  G  E  A  V  F  P  E  G  DM GV WG  R  R  L  D  T
ctggtgagggagctgagcctggcgtggtgtaggtcgaaggggagcccagctgcgccgtg
 L  V  R  E  L  S  L  A  W  C  R  S  K  G  E  P  S  C  A  V
agggaggacctcctcgcggagcaggccaaggcggctgaggtgatcgagaggtcagacggc
 R  E  D  L  L  A  E  Q  A  K  A  A  E  V  I  E  R  S  D  G
gtgatcctcctcaaggcgccgacgggcttcggaaagaccgagatctggacagcgccgttt
 V  I  L  L  K  A  P  T  G  F  G  K  T  E  I  W  T  A  P  F
ttcgcccagtggctcaggggggagtggttcgcgcccaggatgtacgtcgtcgagcccatg
 F  A  Q  W  L  R  G  E  W  F  A  P  R  M  Y  V  V  E  P  M
cacgccctccttaggcagatgaagaggaggatggaggtctacgcccaagccgtccaaggc
 H  A  L  L  R  Q  M  K  R  R  M  E  V  Y  A  Q  A  V  Q  G
ctggggctcccgaggctgaacgtcgctgaggatcacggcgaggtggcgaagcccctcttc
 L  G  L  P  R  L  N  V  A  E  D  H  G  E  V  A  K  P  L  F
ctatacgggggcacatagtcctcaccacagtcgactcgctggcctacggctacctcgcg
 L  Y  G  G  H  I  V  L  T  T  V  D  S  L  A  Y  G  Y  L  A
aggcgggtacagaggtggcgggaggagggcgtggagaggggaaggtacagcatgcccgcg
 R  R  V  Q  R  W  R  E  E  G  V  E  R  G  R  Y  S  M  P  A
ggcctactcgcaagcgcctacatcgtcctagacgaggcccacctcatacaagacgaggcc
 G  L  L  A  S  A  Y  I  V  L  D  E  A  H  L  I  Q  D  E  A
tacctaggccccagggtgctggggaagatagtctgcgatctggcctccgccggagccaag
 Y  L  G  P  R  V  L  G  K  I  V  C  D  L  A  S  A  G  A  K
gtcgtcatatccaccgccacggtccccgagacattcctcaagcacatcccgtgcctcggc
 V  V  I  S  T  A  T  V  P  E  T  F  L  K  H  I  P  C  L  G
```

```
gggaggctgacgcttgggtccggcactgtcagaagaaacgtaaaggtggagaggaggaag
 G  R  L  T  L  G  S  G  T  V  R  R  N  V  K  V  E  R  R  K
ggggtcctcaaggcggaggaaatcgaatgcggcaagcccaccatcgtcattgtgaacact
 G  V  L  K  A  E  E  I  E  C  G  K  P  T  I  V  I  V  N  T
atagagagggcccggcgcatctacaaacaggtgaggtgcgggaagaaggccgtggtgcac
 I  E  R  A  R  R  I  Y  K  Q  V  R  C  G  K  K  A  V  V  H
tcgttgatgaggagggaggacaaggagaggcagctgagcaaggtgctggcggacggaaag
 S  L  M  R  R  E  D  K  E  R  Q  L  S  K  V  L  A  D  G  K
gtggccgaggacgctgtgctgattgggacccaggcgcttgaggtggggctcgacttctcc
 V  A  E  D  A  V  L  I  G  T  Q  A  L  E  V  G  L  D  F  S
aacctaagggcgctctacactgagacggcccccgtagacgcgctgatccagcgcattggg
 N  L  R  A  L  Y  T  E  T  A  P  V  D  A  L  I  Q  R  I  G
agggtggggagagacggggataaggcggaggcctacatctacgaggccgagggcgatgcc
 R  V  G  R  D  G  D  K  A  E  A  Y  I  Y  E  A  E  G  D  A
ccctacccgcagaccctcatgcgggccacgcgcgaagcccttgagaaggagctactggga
 P  Y  P  Q  T  L  M  R  A  T  R  E  A  L  E  K  E  L  L  G
ggcgccgcgctgacgtcttgggaagacgcacagagggccgtggacaaagtgtacaacgag
 G  A  A  L  T  S  W  E  D  A  Q  R  A  V  D  K  V  Y  N  E
aaggccgtggaggagctcatgacgaggggctggcgtggtacggccaagcgctcggctac
 K  A  V  E  E  L  M  T  R  G  L  A  W  Y  G  Q  A  L  G  Y
ctgcaggagctgtccctcttctcctacccgcccagaggagaggtgaggataaggccctcc
 L  Q  E  L  S  L  F  S  Y  P  P  R  G  E  V  R  I  R  P  S
aactacatcacgcttgttattgctgacgtgaagcaggatggggataaggggcggtacatc
 N  Y  I  T  L  V  I  A  D  V  K  Q  D  G  D  K  G  R  Y  I
acggaggacgacgtggagagaggcgcgatgaagatgagctacacctccagagacgacccc
 T  E  D  D  V  E  R  G  A  M  K  M  S  Y  T  S  R  D  D  P
aggattaacgccctgctccagaaggtatccaccgcatatacggtgaggggggttgccacg
 R  I  N  A  L  L  Q  K  V  S  T  A  Y  T  V  R  G  V  A  T
gcgaaggacgagacccgctactacctaagcgagctccggaggagctgggatggggttgag
 A  K  D  E  T  R  Y  Y  L  S  E  L  R  R  S  W  D  G  V  E
gtagttgtggtggacaggagggacgtggaggagctgtacgacgaggcggggctcgacgtg
 V  V  V  V  D  R  R  D  V  E  E  L  Y  D  E  A  G  L  D  V
gcgcagctcagcggcggaggccagaagaggaagggaaggggcggaggaggcgatgagc
 A  Q  L  S  G  G  G  Q  K  R  K  G  R  G  R  R  R  R  M  S
tgctgcgcctacttccgggggagggactgcctccagacctacgaagaccacatcacgcag
 C  C  A  Y  F  R  G  R  D  C  L  Q  T  Y  E  D  H  I  T  Q
gctctggaggcgtgtgaaaggctgaggccctactacggccgctggatggagaaggcattt
 A  L  E  A  C  E  R  L  R  P  Y  Y  G  R  W  M  E  K  A  F
gggaccgccgacgccgccgccctagccgtggagttccacgacctaggcaagctcgccaag
 G  T  A  D  A  A  A  L  A  V  E  F  H  D  L  G  K  L  A  K
gcttacatagccggcaaccgcgccagatataggcacgaggtgttgggggcctactttgcg
 A  Y  I  A  G  N  R  A  R  Y  R  H  E  V  L  G  A  Y  F  A
ttgaagaccctacagacggaggcgagatactacgtggcggcggcggtcgcgctccaccac
 L  K  T  L  Q  T  E  A  R  Y  Y  V  A  A  A  V  A  L  H  H
gagcccatgatcctcgccgcctacgccggcgagctcggcgagagggctatccacgtctcc
 E  P  M  I  L  A  A  Y  A  G  E  L  G  E  R  A  I  H  V  S
acgctgagggccatgctgaaggactccgacctgagcctcggttgcacccccaactacagc
 T  L  R  A  M  L  K  D  S  D  L  S  L  G  C  T  P  N  Y  S
tacaggcctgaggtagccgcggcgctgagggagtgggcaagcagaccgcccaccgccgac
 Y  R  P  E  V  A  A  A  L  R  E  W  A  S  R  P  P  T  A  D
gacgtggccgacgccttccaggagctcgccgtgtatcttagcggcggcgcccccgaggag
 D  V  A  D  A  F  Q  E  L  A  V  Y  L  S  G  G  A  P  E  E
```

```
gcgagggtcaaacggctgagggtggccgggctactacacgtgctgaccgtatgcgacaac
 A  R  V  K  R  L  R  V  A  G  L  L  H  V  L  T  V  C  D  N
tggggagcgaggggccgtccaggagagggaaccttcattagcagatacatgacagtgagc
 W  G  A  R  G  R  P  G  E  G  T  F  I  S  R  Y  M  T  V  S
gagctggggctatgacccagtgcccccaggagatgatcacgccgggccacggcattata
 E  L  G  L  M  T  Q  C  P  Q  E  M  I  T  P  G  H  G  I  I
gccgacgccctcatcatgcacggcctcatcaagatcctccacatcgaggggaagatggag
 A  D  A  L  I  M  H  G  L  I  K  I  L  H  I  E  G  K  M  E
gggtgggcagagcggcgaggcgagagattcgtcgtatgcgccgagaggccggatctgaac
 G  W  A  E  R  R  G  E  R  F  V  V  C  A  E  R  P  D  L  N
tggatagaccagtgggagcccatggagctcctcgagatcgccgcggagttcaaggcgaag
 W  I  D  Q  W  E  P  M  E  L  L  E  I  A  A  E  F  K  A  K
ggcaaggatagggcatcggaagaggaggagcagaggccctacttcggcttggtgcacgac
 G  K  D  R  A  S  E  E  E  E  Q  R  P  Y  F  G  L  V  H  D
ctttccagatccacagtggacccgggcaacttctccagctggatctccgacgtcagagag
 L  S  R  S  T  V  D  P  G  N  F  S  S  W  I  S  D  V  R  E
gcgctgtatgcgctcaaggcggacttcgacctctccgaggagcacaaggagaagaggggg
 A  L  Y  A  L  K  A  D  F  D  L  S  E  E  H  K  E  K  R  G
gagggccgcgcccgctcaaagaaacgcgggtactacacgctctacctccccctctcgggg
 E  G  R  A  R  S  K  K  R  G  Y  Y  T  L  Y  L  P  L  S  G
gtctacggcaaatacgtcgtggagaactacggcataaggcagtcccagtacgcggtctgc
 V  Y  G  K  Y  V  V  E  N  Y  G  I  R  Q  S  Q  Y  A  V  C
gccacgtgcttcgccctcagcaccctgggctacatctacggcaccgtaaaggcgcgggtg
 A  T  C  F  A  L  S  T  L  G  Y  I  Y  G  T  V  K  A  R  V
gagcggaggtcctcctccggcgggggacacgacgtcttcaacctcacgtttatcccaagg
 E  R  R  S  S  S  G  G  G  H  D  V  F  N  L  T  F  I  P  R
gagaggacgtcgctgaggtccctcatggcgcttcagaggatggcaggtctcgtagagatg
 E  R  T  S  L  R  S  L  M  A  L  Q  R  M  A  G  L  V  E  M
cggccgggggatctaaacgagctgggggcagtggtgtacatgctgtcggtcggcgagacc
 R  P  G  D  L  N  E  L  G  A  V  V  Y  M  L  S  V  G  E  T
atatacgcggtgagagagggcgtggacatcctcgtctgggtcacccagagggccggcaac
 I  Y  A  V  R  E  G  V  D  I  L  V  W  V  T  Q  R  A  G  N
ttccagaggaccgtcggggtcaacatattcagaggcgacagactgttggaggccatcgca
 F  Q  R  T  V  G  V  N  I  F  R  G  D  R  L  L  E  A  I  A
gagatcaagtacagagcccccagttggcccaagatagcgaggcagctcggcagctcccta
 E  I  K  Y  R  A  P  S  W  P  K  I  A  R  Q  L  G  S  S  L
aacgtcctaggcgagtacctcgccttcggcggcgacgtgtaccacgtcattaggtcggtg
 N  V  L  G  E  Y  L  A  F  G  G  D  V  Y  H  V  I  R  S  V
atggccgacttggggaggaaaggggggaagttggccgggctggaaggtctcgcagaagcc
 M  A  D  L  G  R  K  G  G  K  L  A  G  L  E  G  L  A  E  A
cttaaaaagataggctagaacaaacccgtgttcaaattcgagttctgccc
 L  K  K  I  G  -
```

Danksagung

Besonders danken möchte ich meinem Doktorvater Prof. Dr. Reinhard Hensel für seine Unterstützung und Hilfe, seine Diskussionsbereitschaft, den wissenschaftlichen Freiraum, sowie für die Bereitstellung des Arbeitsplatzes.

Dr. Britta Tjaden danke ich sehr für ihre Projektvorbereitungen, Ideen und vielen Hilfestellungen im Laboralltag.
Für die technische Assistenz und tatkräftige Unterstützung im Labor danke ich herzlichst Hildegard Eling, Sabine Effenberger und Thomas Knura.

Ein großes Dankeschön gilt allen ehemaligen und jetzigen Mitgliedern der Arbeitsgruppe Mikrobiologie I, der Universität Duisburg-Essen: Karin Bendt, Beatrix Bialek, Roland Diaz-Bone, Britta Huber, Annika Mathias, Dr. Jörg Meyer, Dr. Klaus Michalke, Dr. Patrick Moritz, Kerstin Neuhaus, Dominik Pieper, Annette Schmidt, Veronika Schneider, Andreas Tessarek, Frank Thomas, Dr. Britta Tjaden, Stephanie Vorwerk und Oliver Würfel für die sehr schöne Arbeitsatmosphäre, die große Hilfsbereitschaft, sowie die unzähligen lustigen Momente.
Danken möchte ich auch einigen ehemaligen und jetzigen Mitgliedern der Arbeitsgruppe Molekulare Enzymtechnologie und Biochemie von Prof. Dr. B. Siebers: Dr. Hatim Ahmed, Patrick Haferkamp, Anna Hagemann, Theresa Kouril, Dr. Jeannette Marrero, Kohei Matsubara und Dr. Melanie Zaparty für viele lustige Treffen neben der Laborzeit und Unterstützung bezüglich der Northern Blot Versuche.
Danken möchte ich auch allen jetzigen und ehemaligen Mitgliedern und Stipendiaten des Graduiertenkollegs „Transkription, Chromatinstruktur und DNA Reparatur in Entwicklung und Differenzierung“ für unterhaltsame und lehrreiche Jahrestagungen und für die kritische Diskussionsbereitschaft in zahlreichen Seminaren und Veranstaltungen.
Für intensives und kritisches Korrekturlesen möchte ich besonders Dr. Melanie Zaparty und Stephanie Vorwerk danken.

Elisabeth, meiner Familie und meinen Freunden danke ich ganz besonders für ihr Verständnis, für ihre Hilfe und das sie stets an mich geglaubt haben.

LEBENSLAUF

Persönliche Daten

Name	André Plagens
Anschrift	Schongauerstr. 10 45147 Essen
Geburtsdatum	24. Oktober 1980
Geburtsort	Bottrop
Staatsangehörigkeit	Deutsch
Eltern	Hans-Jürgen Plagens Gabriele Plagens, geb. Nowak

Schulische Ausbildung

1987 – 1991	Droste-Hülshoff Grundschule, Bottrop
1991 – 2000	Josef-Albers Gymnasium, Bottrop

Akademische Ausbildung

2000 – 2006	Studium der Umweltwissenschaften an der Universität Duisburg-Essen
Abschluss	Diplom (Mai 2006; Mikrobiologie, Meeresbiologie, Umweltchemie, Angewandte Limnologie, Biotop- und Artenschutz)
seit Oktober 2006	wissenschaftliche Mitarbeiter in der Arbeitsgruppe Mikrobiologie I, Prof. Dr. R. Hensel und Stipendiat im DFG Graduiertenkolleg „Transkription, Chromatinstruktur und DNA Reparatur in Entwicklung und Differenzierung“, Universität Duisburg-Essen

Diplomarbeit

Zeitraum	März 2005 bis Mai 2006
Betreuer	Prof. Dr. Reinhard Hensel, Universität Duisburg-Essen
Thema	„Enzymatische Charakterisierung der rekombinant exprimierten Proteine Phosphoenolpyruvat Synthetase und Pyruvat, Phosphat Dikinase aus *Thermoproteus tenax* “

Promotionsarbeit

Zeitraum	Oktober 2006 bis Februar 2010
Betreuer	Prof. Dr. Reinhard Hensel, Universität Duisburg-Essen
Thema	“Characterisation of the CRISPR/Cas system of the hyperthermophilic Archaeum *Thermoproteus tenax”*

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 7 der Promotionsordnung der Math.-Nat.-Fakultäten zur Erlangung des Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema “Characterisation of the CRISPR/Cas system of the hyperthermophilic Archaeum *Thermoproteus tenax”* zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Herrn André Plagens befürworte.

Essen, den 16.02.2010 _____

Prof. Dr. Reinhard Hensel

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 6 der Promotionsordnung der Math.-Nat.-Fakultäten zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient habe.

Essen, den 16.02.2010 _____

André Plagens

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. 2, Nr. 8 der Promotionsordnung der Math.-Nat.-Fakultäten zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den 16.02.2010 _____

André Plagens